Chương 6

MỞ RỘNG MÔ HÌNH HỒI QUY TUYẾN TÍNH HAI BIẾN

Một số khía cạnh của phân tích hồi quy tuyến tính có thể được dễ dàng trình bày trong khuôn khổ mô hình hồi quy tuyến tính hai biến như ta đã thảo luận từ đầu cho tới nay. Trước hết, ta xem xét trường hợp **hồi quy qua gốc tọa độ**, tức là, một tình thế mà số hạng tung độ gốc, β_1, không có trong mô hình. Sau đó, ta sẽ xem xét vấn đề **đơn vị đo**, tức là, biến Y và X được đo như thế nào và việc thay đổi đơn vị đo có tác động tới các kết quả hồi quy hay không. Sau cùng, ta phân tích vấn đề **dạng hàm số** của mô hình hồi quy tuyến tính. Cho tới nay, ta đã phân tích các mô hình tuyến tính theo với các thông số và theo các biến số. Nhưng nhớ lại rằng lý thuyết hồi quy xây dựng trong các chương trước chỉ yêu cầu phải tuyến tính theo các thông số; các biến số có thể tuyến tính hay phi tuyến trong mô hình. Bằng cách xem xét mô hình tuyến tính theo các thông số nhưng không nhất thiết tuyến tính theo các biến số, ta sẽ chỉ ra trong chương này làm sao các mô hình hai biến có thể giải quyết một số vấn đề thực tế thú vị.

Sau khi nắm bắt được các ý tưởng trong chương này, việc mở rộng ra các mô hình hồi quy bội là điều khá dễ dàng, như ta sẽ trình bày trong Chương 7 và 8.

6.1. HỒI QUY QUA GỐC TỌA ĐỘ

Có những trường hợp mà hàm hồi quy tổng thể (PRF) hai biến có dạng sau:

$$Y_i = \beta_2 X_i + u_i \tag{6.1.1}$$

Trong mô hình này, tung độ gốc không có hay bằng 0. Do vậy, dạng mô hình này có tên là **hồi quy qua gốc tọa độ**.

Để minh họa, hãy xem xét Mô hình xác định giá tài sản vốn (Capital Asset Pricing Model - CAPM) trong lý thuyết cơ cấu đầu tư chứng khoán hiện đại. Trong dạng rủi ro - thưởng kim, có thể được thể hiện như sau:[1]

$$(ER_i - r_f) = \beta_i(ER_m - r_f) \tag{6.1.2}$$

với ER_i = suất sinh lợi kỳ vọng của chứng khoán i

ER_m = suất sinh lợi kỳ vọng trung bình của cơ cấu chứng khoán thị trường ví dụ như được đại diện bởi chỉ số cổ phiếu tổng hợp S&P 500.

r_f = suất sinh lợi không có rủi ro, ví dụ lãi suất của tín phiếu kho bạc 90 ngày.

β_i = hệ số Bê ta, một đại lượng đo rủi ro có tính hệ thống, nghĩa là rủi ro không thể bị loại bỏ bằng cách đa dạng hóa chứng khoán. Đồng thời nó cũng là đại lượng đo

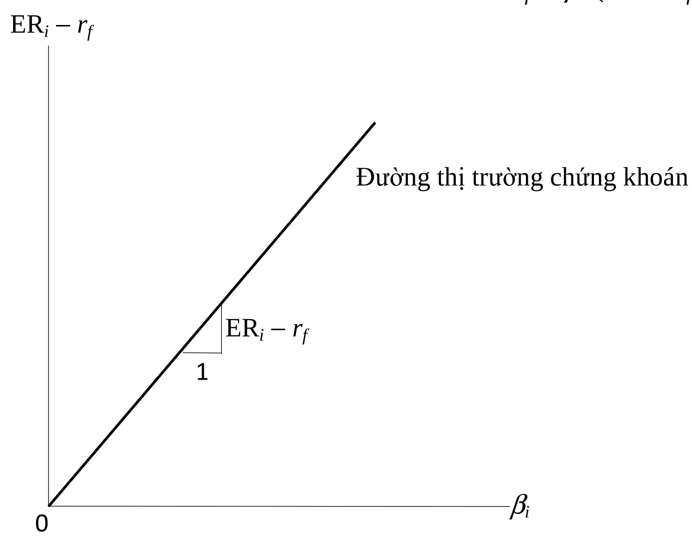
[1] Xem Haim Levy & Marshal Sarnat, *Portfolio and Investment Selection: Theory and Practice* (Lựa chọn cơ cấu chứng khoán và đầu tư: Lý thuyết và thực hành), Prentice-Hall International, Englewood Cliffs, N.J., 1984, Chương 14.

chuyển dịch của suất sinh lợi của chứng khoán *i* theo thị trường. $\beta_i > 1$ có nghĩa là chứng khoán hay thay đổi hay năng động, trái lại $\beta_i < 1$ là chứng khoán phòng thủ. (*Lưu ý*: Đừng nhầm lẫn β_i ở đây với hệ số độ dốc của hồi quy hai biến, β_2).

Nếu các thị trường vốn hoạt động hiệu quả thì CAPM mặc định rằng mức thưởng kim rủi ro kỳ vọng của chứng khoán *i* (= $ER_i - r_f$) bằng với hệ số β nhân với mức thưởng kim rủi ro kỳ vọng của thị trường (= $ER_m - r_f$). Nếu CAPM thỏa mãn, ta có tình trạng như trong Hình 6.1. Đường thẳng trong hình được gọi là **đường thị trường chứng khoán** (SML).

Đối với các mục đích thực nghiệm, (6.1.2) thường được biểu diễn là:

$$R_i - r_f = \beta_1(R_m - r_f) + u_i \qquad (6.1.3)$$



HÌNH 6.1
Rủi ro có tính hệ thống

hay

$$R_i - r_f = \alpha_i + \beta_i(R_m - r_f) + u_i \qquad (6.1.4)$$

Mô hình sau được gọi là **Mô hình thị trường**.[2] Nếu CAPM thỏa mãn, α_i được kỳ vọng là sẽ bằng 0. (Xem hình 6.2).

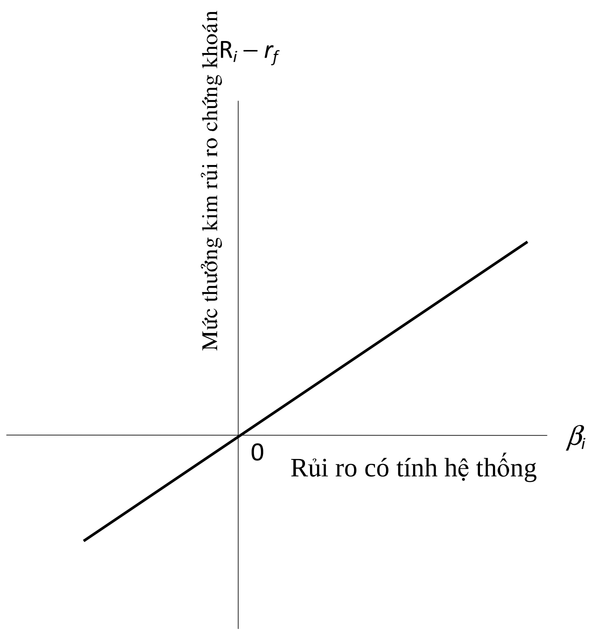
Trong khi chuyển sang phần khác, lưu ý rằng trong (6.1.4) biến phụ thuộc, *Y*, là ($R_i - r_f$) và biến giải thích, *X*, là β_i, hệ số về tính không ổn định, chứ không phải là ($R_m - r_f$). Do vậy, để chạy hồi quy (6.1.4), trước hết ta phải ước lượng β_i. β_i thường được tính từ **đường đặc tính**, như mô tả trong bài tập 5.5. (Về chi tiết, xem bài tập 8.34).

[2] Ví dụ, xem Diana R. Harrington, *Modern Portfolio Theory and the Capital Asset Pricing Model: A User's Guide* (Lý thuyết đầu tư chứng khoán hiện đại và mô hình định giá tài sản vốn: Sách hướng dẫn người sử dụng), Prentice-Hall, Englewood Cliffs, N.J., 1983, trang 71.

Như ví dụ này minh họa, đôi khi trong lý thuyết cơ bản phát biểu rằng tung độ gốc sẽ không có trong mô hình. Các ví dụ khác về các mô hình có tung độ gốc bằng 0 có thể thích hợp là giả thiết về thu nhập thường xuyên của Milton Friedman, trong đó phát biểu rằng tiêu dùng thường xuyên tỷ lệ thuận với thu nhập thường xuyên; lý thuyết phân tích chi phí, trong đó mặc định rằng chi phí sản xuất khả biến tỷ lệ thuận với sản lượng; và một số dạng của lý thuyết tiền tệ, trong đó phát biểu rằng tốc độ thay đổi giá cả (nghĩa là tỷ lệ lạm phát) tỷ lệ thuận với tốc độ thay đổi lượng cung tiền.

Làm sao chúng ta ước lượng các mô hình như (6.1.1) và mô hình này đặt ra các vấn đề đặc biệt nào? Để trả lời các câu hỏi này, trước hết hãy viết hàm hồi quy mẫu (SRF) của (6.1.1), cụ thể là,

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i \tag{6.1.5}$$



HÌNH 6.2
Mô hình thị trường của lý thuyết cơ cấu đầu tư chứng khoán (giả sử là $\alpha_i = 0$).

Bây giờ áp dụng phương pháp OLS đối với (6.1.5), ta có các công thức sau cho $\hat{\beta}_2$ và phương sai của nó (các chứng minh được trình bày trong Phụ lục 6A, Mục 6A.1):

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \tag{6.1.6}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2} \tag{6.1.7}$$

với σ^2 được ước lượng bởi

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1} \tag{6.1.8}$$

So sánh các công thức này với các công thức khi có số hạng tung độ gốc trong mô hình:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{3.1.6}$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{3.3.1}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{3.3.5}$$

Các sự khác biệt giữa hai tập hợp công thức rất rõ ràng: trong mô hình không có số hạng tung độ gốc, ta sử dụng tổng bình phương và tích chéo **thô** nhưng trong mô hình có tung độ gốc, ta sử dụng tổng bình phương và tích chéo hiệu chỉnh (từ giá trị trung bình). Thứ hai, số bậc tự do để tính $\hat{\sigma}^2$ là (n – 1) trong trường hợp thứ nhất và (n – 2) trong trường hợp thứ hai. (Tại sao?)

Mặc dù mô hình không có tung độ gốc hay tung độ gốc bằng 0 có thể thích hợp trong một số trường hợp, có một số đặc điểm của mô hình này mà ta cần phải lưu ý. Thứ nhất, $\sum \hat{u}_i$, luôn bằng 0 trong mô hình có tung độ gốc (mô hình quy ước), nhưng không cần phải bằng 0 trong trường hợp không có tung độ gốc. Nói ngắn gọn, $\sum \hat{u}_i$ không nhất thiết phải bằng 0 đối với hồi quy qua gốc tọa độ. Thứ hai, r^2, hệ số xác định giới thiệu trong Chương 3, luôn không âm đối với mô hình quy ước, nhưng có thể trong một số trường hợp trở nên âm trong mô hình không có tung độ gốc! Kết quả bất thường này phát sinh do r^2 trình bày trong Chương 3 giả sử một cách rõ ràng rằng mô hình chứa tung độ gốc. Do vậy, r^2 tính theo cách quy ước có thể không thích hợp cho các mô hình hồi quy qua gốc tọa độ.[3]

r^2 đối với mô hình hồi quy qua gốc tọa độ

Như vừa lưu ý, và thảo luận sâu hơn trong Phụ lục 6A, Mục 6A.1, r^2 quy ước trong Chương 3 không thích hợp cho hồi quy không có tung độ gốc. Nhưng ta có thể tính cái gọi là ***r*²** **thô** cho các mô hình như vậy theo công thức sau:

$$r^2 \text{ thô} = \frac{\sum (X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2} \tag{6.1.9}$$

Lưu ý: Đây là các tổng bình phương và tích chéo thô (nghĩa là không được hiệu chỉnh theo trung bình).

[3] Về các phần thảo luận thêm, xem Dennis J. Aigner, *Basic Econometrics* (Kinh lượng cơ bản), Prentice-Hall, Englewood Cliffs, N.J., 1971, trang 85-88.

Mặc dù hệ số r^2 thô này thỏa mãn quan hệ $0 < r^2 < 1$, nó không thể so sánh trực tiếp với giá trị r^2 quy ước. Vì lý do này, một số tác giả không báo cáo giá trị r^2 đối với các mô hình hồi quy có tung độ gốc bằng 0.

Do các đặc điểm cụ thể của mô hình này, ta cần rất cẩn thận khi sử dụng mô hình hồi quy có gốc tọa độ bằng 0. *Trừ khi có một tiên nghiệm rất mạnh*, ta cần phải sử dụng mô hình quy ước, có tung độ gốc. Điều này có một lợi thế kép. Thứ nhất, nếu số hạng tung độ gốc được đưa vào mô hình nhưng nó trở nên không có ý nghĩa về mặt thống kê (nghĩa là, bằng 0 về mặt thống kê), đối với tất cả các mục đích thực tế, ta có một hồi quy qua gốc tọa độ.[4] Thứ hai, và quan trọng hơn, nếu thật sự có tung độ gốc nhưng ta khẳng đnh rằng hồi qui gốc tọa độ, ta sẽ phạm **sai số đặc trưng**, và như vậy vi phạm Giả thiết 9 của mô hình hồi quy tuyến tính cổ điển.

Ví dụ minh họa: Đường đặc tính của lý thuyết cơ cấu đầu tư chứng khoán

Bảng 6.1 cung cấp số liệu về suất sinh lợi hàng năm (%) của Afuture Fund, một quỹ hỗ tương có mục tiêu đầu tư cơ bản là tối đa lợi nhuận từ tăng giá trị vốn, và suất sinh lợi trung bình của cơ cấu chứng khoán thị trường, tính bởi Chỉ số Fisher, trong giai đoạn 1971-1980.

Trong bài tập 5.5, ta đã giới thiệu *đường đặc tính* của phân tích đầu tư. Đường này có thể được biểu diễn như sau:

$$Y_i = \alpha_i + \beta_i X_i + u_I \tag{6.1.10}$$

với Y_i = suất sinh lợi hàng năm (%) của Afuture Fund
X_i = suất sinh lợi hàng năm (%) của cơ cấu chứng khoán thị trường
β_i = hệ số độ dốc, cũng được gọi là hệ số **Bê ta** trong lý thuyết cơ cấu đầu tư chứng khoán, và
α_i = tung độ gốc

Trong lý thuyết, các nhà nghiên cứu không đạt được một sự đồng tình về giá trị có trước của α_i. Một số kết quả thực nghiệm đã cho thấy α_i dương và có ý nghĩa thống kê và một số khác lại cho thấy nó không khác 0 một cách có ý nghĩa về thống kê; trong trường hợp sau ta có thể viết mô hình dưới dạng:

$$Y_i = \beta_i X_i + u_i \tag{6.1.11}$$

tức là, một hồi quy qua gốc tọa độ.

[4] Henri Theil chỉ ra rằng nếu tung độ gốc thật sự không có, hệ số độ dốc có thể được ước lượng với độ chính xác lớn hơn rất nhiều so với trường hợp có tung độ gốc. Xem *Introduction to Econometrics* (Giới thiệu Kinh tế lượng) của Henri Theil, Prentice-Hall, Englewood Cliffs, N.J., 1978, trang 76. Xem đồng thời vị dụ số trong phần sau.

BẢNG 6.1
Suất sinh lợi trung bình của Afuture Fund và của Chỉ số Fisher (cơ cấu chứng khoán thị trường), 1971-1980

Năm	**Suất sinh lợi của Afuture Fund (%)** ***Y***	**Suất sinh lợi dựa trên Chỉ số Fisher (%)** ***X***
1971	37,5	19,5
1972	19,2	8,5
1973	-35,2	-29,3
1974	-42,0	-26,5
1975	63,7	61,9
1976	19,3	45,5
1977	3,6	9,5
1978	20,0	14,0
1979	40,3	35,3
1980	37,5	31,0

Nguồn: Haim Levy & Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice* (Lựa chọn cơ cấu chứng khoán và đầu tư: Lý thuyết và thực hành), Prentice-Hall, Englewood Cliffs, N.J., 1984, trang 730 & 738. Các số liệu này được thư thập bởi các tác giả từ Weisenberg Investment Service, *Investment Companies*, lần xuất bản 1981.

Nếu quyết định sử dụng mô hình (6.1.1), ta có các kết quả hồi quy sau (xem kết quả in ra của SAS trong Phụ lục 6A, Mục 6A.2):

$$
\begin{aligned}
\hat{Y}_i &= 1{,}0899X_i \\
&\quad (0{,}1916) \qquad r^2 \text{ thô} = 0{,}7825 \\
t &= \ (5{,}6884)
\end{aligned}
\tag{6.1.12}
$$

kết quả này cho thấy β_i lớn hơn 0 về ý nghĩa thống kê. Sự giải thích là 1% tăng của suất sinh lợi thị trường sẽ làm tăng trung bình 1,09% suất sinh lợi của Afuture Fund.

Làm sao chúng ta có thể chắc chắn rằng mô hình (6.1.11), chứ không phải (6.1.10) là thích hợp, đặc biệt là trong trường hợp không có một tiên nghiệm mạnh trong giả thiết là α_i thật sự bằng 0? Điều này có thể được kiểm tra bằng cách chạy hồi quy (6.1.10). Sử dụng số liệu trong Bảng 6.1, ta có các kết quả sau:

$$
\begin{aligned}
\hat{Y}_i &= \ 1{,}2797 \quad + \quad 1{,}0691X_i \\
&\quad (7{,}6886) \qquad (0{,}2383) \\
t &= \ (0{,}1664) \qquad (4{,}4860) \qquad r^2 = 0{,}7155
\end{aligned}
\tag{6.1.13}
$$

Lưu ý: Các giá trị r^2 của (6.1.12) và (6.1.13) *không* trực tiếp so sánh với nhau. Từ những kết quả này, ta không thể bác bỏ giả thiết cho rằng giá trị đúng của tung độ gốc bằng 0, do vậy xác nhận cho việc sử dụng (6.1.1), tức là hồi quy qua gốc tọa độ.

Trong khi chuyển sang phần khác, lưu ý rằng không có sự khác nhau nhiều giữa các kết quả của (6.1.12) và (6.1.13), mặc dù sai số chuẩn ước lượng của $\hat{\beta}_2$ hơi nhỏ hơn trong mô hình hồi quy qua gốc tọa độ, và do vậy hỗ trợ lập luận của Theil trong chú thích 4 cho rằng nếu α_i thật sự bằng 0, hệ số độ dốc có thể được tính với độ chính xác cao hơn: sử dụng số liệu trong Bảng 6.1 và các kết quả hồi quy, người đọc có thể dễ dàng chứng minh rằng khoảng tin cậy 95% đối với hệ số độ dốc của mô hình hồi quy qua gốc tọa độ là (0,6566, 1,5232), trong khi đối với mô hình (6.1.13), khoảng tin cậy này là (0,5195, 1,6186); tức là, khoảng tin cậy trước hẹp hơn khoảng tin cậy sau.

6.2 TỶ LỆ VÀ ĐƠN VỊ ĐO

Để nắm bắt các ý tưởng phát triển trong mục này, hãy xem xét số liệu trong Bảng 6.2. Bảng này cung cấp số liệu về tổng đầu tư tư nhân nội địa (GPDI) và tổng sản phẩm quốc dân (GNP) theo giá đô la năm 1972 trong giai đoạn 1974-1983. Cột (1) trình bày số liệu về GPDI tính theo tỷ USD, trong khi cột (2) trình bày GPDI tính theo triệu USD. Cột (3) và (4) trình bày số liệu GNP tương ứng theo tỷ và triệu USD.

Giả sử trong hồi quy của GPDI đối với GNP, một nhà nghiên cứu sử dụng số liệu tính theo tỷ USD nhưng một người khác lại sử dụng những biến này tính theo triệu USD. Các kết quả hồi quy trong hai trường hợp có giống nhau không? Nếu không, chúng ta phải sử dụng các kết quả nào? Nói một cách ngắn gọn, các đơn vị đo của biến *Y* và *X* có tạo nên một sự khác biệt nào không trong các kết quả hồi quy? Nếu có thì đâu là cách thức đúng đắn để lựa chọn đơn vị đo trong phân tích hồi quy?

Để trả lời các câu hỏi này, hãy tiến hành một cách hệ thống. Đặt

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{6.2.1}$$

với *Y* = GPDI và *X* = GNP. Định nghĩa:

$$Y_i^* = w_1 Y_i \tag{6.2.2}$$

$$X_i^* = w_2 X_i \tag{6.2.3}$$

với w_1 và w_2 là các hằng số, gọi là các **hệ số tỷ lệ**; w_1 có thể bằng hoặc khác w_2.

BẢNG 6.2 Tổng đầu tư tư nhân nội địa (GPDI) và tổng sản phẩm quốc dân (GNP) theo giá 1972, Hoa Kỳ, 1974-1983

Năm	GPDI (tỷ USD, giá 1972) (1)	GPDI (triệu USD, giá 1972) (2)	GNP (tỷ USD, giá 1972) (3)	GNP (triệu USD, giá 1972) (4)
1974	195,5	195.500	2146,3	2.146.300
1975	154,8	154.800	1231,6	1.231.600
1976	184,5	184.500	1298,2	1.298.200
1977	214,2	214.200	1369,7	1.369.700
1978	236,7	236.700	1438,6	1.438.600

1979	236,3	236.300	1479,4	1.479.400
1980	208,5	208.500	1475,0	1.475.000
1981	230,9	230.900	1512,2	1.512.200
1982	194,3	194.300	1480,0	1.480.000
1983	221,0	221.000	1534,7	1.534.700

Nguồn: Báo cáo Kinh tế của Tổng thống, 1985, trang 234 (cho số liệu tính theo tỷ USD).

Từ (6.2.2) và (6.2.3), rõ ràng là Y_i^* và X_i^* là Y_i và X_i được tính lại theo tỷ lệ khác. Như vậy, nếu Y_i và X_i được tính bằng tỷ USD và ta muốn biểu diễn chúng dưới dạng triệu USD, ta có: $Y_i^* = 1000\, Y_i$ và $X_i^* = 1000\, X_i$; trong trường hợp này $w_1 = w_2 = 1000$.

Xem xét hồi quy sau sử dụng các biến Y_i^* và X_i^*:

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_i^* X_i + \hat{u}_i^* \tag{6.2.4}$$

với $Y_i^* = w_1 Y_i$, $X_i^* = w_2 X_i$, và $\hat{u}_i^* = w_1 \hat{u}_i$. (Tại sao?)

Ta muốn tìm các mối quan hệ giữa các cặp sau:

1. $\hat{\beta}_1$ và $\hat{\beta}_1^*$
2. $\hat{\beta}_2$ và $\hat{\beta}_2^*$
3. var($\hat{\beta}_1$) và var($\hat{\beta}_1^*$)
4. var($\hat{\beta}_2$) và var($\hat{\beta}_2^*$)
5. $\hat{\sigma}^2$ và $\hat{\sigma}^{*2}$
6. r_{xy}^2 và $r_{x^*y^*}^2$

Từ lý thuyết bình phương tối thiểu, ta biết rằng (xem Chương 3):

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \tag{6.2.5}$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{6.2.6}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2 \tag{6.2.7}$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{6.2.8}$$

$$\sigma^2 = \frac{\sum u_i^2}{n-2} \tag{6.2.9}$$

Áp dụng phương pháp OLS cho (6.2.4), ta có tương tự:

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \tag{6.2.10}$$

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} \tag{6.2.11}$$

$$\text{var}(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n\sum x_i^{*2}}\sigma^{*2} \tag{6.2.12}$$

$$\text{var}(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \tag{6.2.13}$$

$$\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{(n-2)} \tag{6.2.14}$$

Từ các kết quả này, ta dễ dàng thiết lập các mối quan hệ giữa hai tập hợp các ước lượng thông số. Tất cả những gì phải làm là nhớ lại những quan hệ định nghĩa: $Y_i^* = w_1 Y_i$ (hay $y_i^* = w_1 y_i$); $X_i^* = w_2 X_i$ (hay $x_i^* = w_2 x_i$); $\hat{u}_i^* = w_1 \hat{u}_i$; $\bar{Y}^* = w_1 \bar{Y}$ và $\bar{X}^* = w_2 \bar{X}$. Sử dụng các định nghĩa này, người đọc có thể dễ dàng chứng minh rằng:

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)\hat{\beta}_2 \tag{6.2.15}$$

$$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \tag{6.2.16}$$

$$\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2 \tag{6.2.17}$$

$$\text{var}(\hat{\beta}_1^*) = w_1^2 \,\text{var}(\hat{\beta}_1) \tag{6.2.18}$$

$$\text{var}\,\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)^2 \text{var}(\hat{\beta}_2) \tag{6.2.19}$$

$$r_{xy}^2 = r_{x^*y^*}^2 \tag{6.2.20}$$

Từ các kết quả trên, ta thấy rõ rằng với các kết quả hồi quy dựa vào một tỷ lệ đo, ta có thể tính các kết quả dựa trên một tỷ lệ khác khi biết được các hệ số tỷ lệ, *w*. Trên thực tế, mặc dù ta phải lựa chọn đơn vị đo một cách hợp lý, rất có ít ý nghĩa trong việc dùng tất cả các con số 0 để biểu diễn các số hàng triệu và hàng tỷ USD.

Từ các kết quả (6.2.15) tới (6.2.20), ta có thể dễ dàng tính một số trường hợp đặc biệt. Ví dụ, nếu $w_1 = w_2$, tức là các hệ số tỷ lệ đồng nhất, hệ số độ dốc và sai số chuẩn của nó không đổi khi chuyển từ tỷ lệ (Y_i, X_i) sang (Y_i^*, X_i^*). Điều này rất rõ ràng về mặt trực giác. Tuy nhiên, tung độ gốc và sai chuẩn của nó đều được nhân lên bởi w_1. Nhưng nếu tỷ lệ *X* không đổi (nghĩa là $w_2 = 1$) và tỷ lệ *Y* thay đổi bởi hệ số w_1, hệ số độ dốc lẫn tung độ gốc và sai số chuẩn tương ứng của chúng đều được nhân lên bởi cùng hệ số w_1. Sau cùng, nếu tỷ lệ *Y* không đổi (nghĩa là $w_1 = 1$) và tỷ lệ *X* thay đổi bởi hệ số w_2, hệ số độ dốc và sai số chuẩn của nó được nhân lợi bởi hệ số ($1/w_2$) nhưng tung độ gốc và sai số chuẩn của nó không đổi.

Tuy nhiên, phải lưu ý rằng việc chuyển đổi từ tỷ lệ (Y_i, X_i) sang (Y_i^*, X_i^*) không tác động tới những tính chất của các ước lượng OLS thảo luận trong các chương trước.

Ví dụ số: Quan hệ giữa GDPI và GNP, Hoa Kỳ, 1974-1983

Để chứng minh các kết quả lý thuyết ở trên, hãy quay lại với ví dụ trong Bảng 6.2 và xem xét các kết quả hồi quy sau. (Các số liệu trong ngoặc là sai số chuẩn ước lượng).

Cả GPDI và GNP tính theo tỷ USD:

$$\begin{aligned}\widehat{\text{GPDI}}_t &= -37{,}0015205 + 0{,}17395\ \text{GNP}_t \\ &\quad (76{,}2611278)\quad (0{,}05406) \\ r^2 &= 0{,}5641\end{aligned} \tag{6.2.21}$$

Cả GPDI và GNP tính theo triệu USD:

$$\begin{aligned}\widehat{\text{GPDI}}_t &= -37001{,}5205 + 0{,}17395\ \text{GNP}_t \\ &\quad (76261{,}1278)\quad (0{,}05406) \\ r^2 &= 0{,}5641\end{aligned} \tag{6.2.22}$$

Lưu ý rằng tung đô gốc cũng như sai số chuẩn của nó là 1000 (nghĩa là w_i = 1000 trong chuyển đổi từ tỷ sang triệu USD) nhân với các giá trị tương ứng trong hồi quy (6.2.21), nhưng hệ số độ dốc cũng như sai số chuẩn của nó không đổi, như theo lý thuyết.

GDPI tính theo tỷ USD và GNP tính theo triệu USD:

$$\begin{aligned}\widehat{\text{GPDI}}_t &= -37{,}0015205 + 0{,}00017395\ \text{GNP}_t \\ &\quad (76{,}2611278)\quad (0{,}00005406) \\ r^2 &= 0{,}5641\end{aligned} \tag{6.2.23}$$

Như dự kiến, hệ số độ dốc cũng như sai số chuẩn của nó là (1/1000) giá trị của nó trong (6.2.21) do chỉ có tỷ lệ của *X* hay GNP được thay đổi.

GDPI tính theo triệu USD và GNP tính theo tỷ USD:

$$\begin{aligned}\widehat{\text{GPDI}}_t &= -37001{,}5205 + 173{,}95\ \text{GNP}_t \\ &\quad (76261{,}1278)\quad (54{,}06) \\ r^2 &= 0{,}5641\end{aligned} \tag{6.2.24}$$

Một lần nữa, lưu ý rằng cả tung độ gốc và hệ số độ dốc cũng như sai số tương ứng của chúng bằng 1000 lần giá trị của chúng trong (6.2.21), theo như các kết quả lý thuyết của chúng ta.

Một vài lời giải thích

Do hệ số độ dốc, $\hat{\beta}_2$, đơn giản là tỷ lệ thay đổi, nó được tính bởi đơn vị của tỷ lệ[5]

$$\frac{\text{Đơn vị của biến phụ thuộc}, Y}{\text{Đơn vị của biến giải thích}, X}$$

[5] Về chi tiết và mở rộng sang hồi quy bội, xem Donald F. Morrison, *Applied Linear Statistical Methods* (Các phương pháp thống kế tuyến tính ứng dụng), Prentice-Hall, Englewood Cliffs, N.J., 1983, trang 72.

Như vậy trong hồi quy (6.2.21), sự giải thích về hệ số độ dốc 0,17395 là nếu GNP thay đổi đi một đơn vị, trong trường hợp này là 1 tỷ USD, tính trung bình GPDI thay đổi đi 0,17395 tỷ USD. Trong hồi quy (6.2.23) nếu GNP thay đổi đi một đơn vị, trong trường hợp này là 1 triệu USD, tính trung bình GPDI thay đổi đi 0,00017395 tỷ USD. Hai kết quả tất nhiên là đồng nhất với nhau trong tác động của GNP tới GPDI; đơn giản là chúng được biểu diễn bởi các đơn vị đo khác nhau.

6.3 CÁC DẠNG HÀM SỐ CỦA NHỮNG MÔ HÌNH HỒI QUY

Như đã lưu ý ở Chương 2, cuốn sách này phân tích chủ yếu các mô hình tuyến tính theo các thông số thống kê, chúng có thể tuyến tính hay không tuyến tính theo các biến số. Trong các mục sau, ta xem xét một số mô hình hồi quy thường được sử dụng có thể phi tuyến theo các biến số nhưng phải là tuyến tính theo các thông số hay có thể được dễ dàng tuyến tính hóa bằng các biến đổi thích hợp của biến số. Cụ thể, ta thảo luận các mô hình hồi quy sau:

1. Mô hình tuyến tính lôgarít
2. Mô hình bán lôgarít (semilog)
3. Mô hình nghịch đảo

Ta thảo luận các điểm đặc biệt của từng mô hình, áp dụng chúng trong trường hợp nào và làm sao ước lượng được chúng. Mỗi mô hình được minh họa bởi những ví dụ phù hợp.

6.4 LÀM THẾ NÀO ĐỂ TÍNH ĐỘ CO GIÃN: MÔ HÌNH TUYẾN TÍNH LÔGARÍT

Xem xét mô hình sau với tên gọi là **mô hình hồi quy mũ**:

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \tag{6.4.1}$$

Phương trình có thể được biểu diễn dưới dạng sau:[6]

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i \tag{6.4.2}$$

với ln = lôgarít tự nhiên (nghĩa là log cơ số *e*, với *e* = 2,718).[7]

Nếu ta viết (6.4.2) dưới dạng:

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \tag{6.4.3}$$

với $\alpha = \ln\beta_1$, mô hình này tuyến tính theo các thông số α và β_2, tuyến tính theo lôgarít của các biến *Y* và *X*. Mô hình có thể được ước lượng bằng hồi quy OLS. Do tính chất tuyến tính này, các mô hình như thế được gọi là mô hình **log-log**, **log kép**, hay **tuyến tính log**.

---

[6] Lưu ý những tính chất của lôgarít: (1) ln(*AB*) = ln(*A*) + ln(*B*), (2) ln(*A*/*B*) = ln(*A*) – ln(*B*), và (3) $\ln(A^k) = k\ln(A)$, giả sử *A* và *B* dương và *k* là số không đổi.

[7] Trên thực tế ta có thể sử dụng lôgarít thập phân, tức là, log cơ số 10. Quan hệ giữa log tự nghiên và log thập phân là: $\ln_e X = 2{,}306\log_{10}X$. Theo quy ước, ln biểu thị lôgarít tự nhiên và log biểu thị lôgarít cơ số 10; do vậy không cần phải viết rõ những ký tự con *e* và 10 này.

Nếu các giả thiết của mô hình hồi quy tuyến tính cổ điển được thỏa mãn, các thông số của (6.4.3) có thể được ước lượng bằng phương pháp OLS thông qua việc đặt:

$$Y_i^* = \alpha + \beta X_i^* + u_i \tag{6.4.4}$$

với Y_i^* = lnY_i và X_i^* = lnX_i. Các ước lượng $\hat{\alpha}$ và $\hat{\beta}_2$ tính được sẽ là các ước lượng tuyến tính không thiên lệch tốt nhất, tương ứng của α và β_2.

Một đặc điểm lý thú của mô hình log-log đã làm nó trở thành thông dụng trong nghiên cứu ứng dụng là hệ số độ dốc β_2 đo **độ co giãn** của *Y* so với *X*, tức là, tỷ lệ phần trăm thay đổi *Y* với một tỷ lệ phần trăm thay đổi (nhỏ) cho trước của *X*.[8] Như vậy, nếu *Y* đại diện cho lượng cầu hàng hóa và *X* đại diện cho giá trung bình của nó, β_2 đo hệ số co giãn giá cả của cầu, một thông số có mối quan tâm kinh tế quan trọng. Nếu mối quan hệ giữa lượng cầu và giá cả được biểu diễn như trong Hình 6.3a, việc chuyển đổi log-kép như trong hình 6.3b sẽ cho ta ước lượng của độ co giãn giá cả ($-\beta_2$).

Lưu ý hai đặc điểm của mô hình tuyến tính logarít: Mô hình giả sử rằng hệ số co giãn giữa *Y* và *X*, β_2, không đổi (tại sao?). Do vậy, mô hình có một tên gọi nữa là **mô hình hệ số co giãn không đổi**.[9] Nói một cách khác, như Hình 6.3b biểu thị, thay đổi của ln*Y* khi ln*X* thay đổi đi một đơn vị (nghĩa là hệ số co giãn, β_2) không đổi, không phụ thuộc vào giá trị tuyệt đối của ln*X* mà ta dựa vào để tính hệ số co giãn. Một đặc điểm nữa của mô hình là mặc dù $\hat{\alpha}$ và $\hat{\beta}_2$ là các ước lượng không thiên lệch của α và β_2, β_1 (thông số trong mô hình gốc) khi ước lượng là $\hat{\beta}_1$ = đối logarit($\hat{\alpha}$), bản thân nó là một ước lượng thiên lệch. Tuy nhiên trong hầu hết các vấn đề khó khăn trong thực tiễn, số hạng tung độ gốc có tầm quan trọng thứ hai, và ta không cần lo lắng tới việc tính ước lượng không thiên lệch của nó.[10]

---

[8] Hệ số co giãn, theo cách viết giải tích, được định nghĩa là $(dY/Y)(dX/X) = [(dY/dX)(X/Y)]$. Những người đọc quen thuộc với cách tính vi phân sẽ nhận thấy ngay rằng β_2 trên thực tế là hệ số co giãn.

Lưu ý kỹ thuật: Người đọc suy luận theo phương pháp giải tích sẽ lưu ý rằng $d(\ln X)/dX = 1/X$ hay $d(\ln X) = dX/X$, tức là, đối với những thay đổi vô cùng nhỏ (lưu ý tới toán tử vi phân *d*) thay đổi của ln*X* bằng với thay đổi tương đối hay tỷ lệ của *X*. Trên thực tế, nếu thay đổi của *X* nhỏ, quan hệ này có thể được viết theo dạng: thay đổi của ln*X* Ñ thay đổi tương đối của *X*, với Ñ nghĩa là gần đúng. Như vậy, với những thay đổi nhỏ,

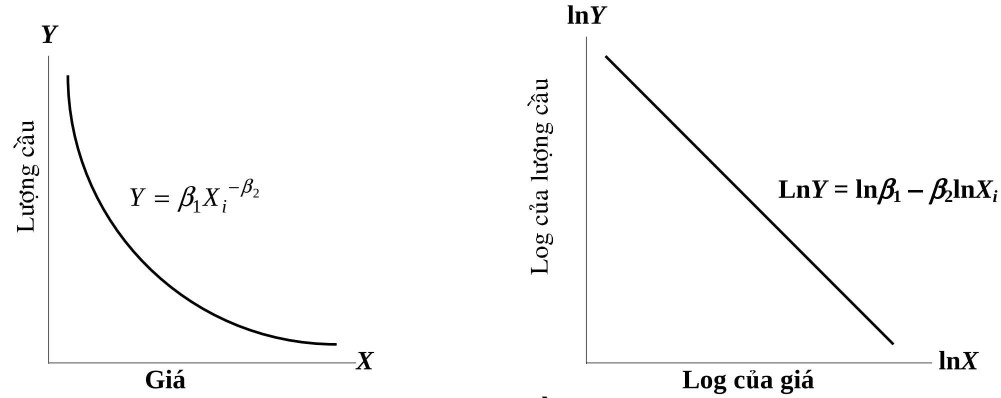
$(\ln X_t - \ln X_{t-1})$ Ñ $(X_t - X_{t-1})/X_{t-1}$ = thay đổi tương đối của *X*

Nhân tiện, người đọc phải lưu ý các thuật ngữ hay gặp phải sau: (1) **thay đổi tuyệt đối**, (2) **thay đổi tương đối** hay **tỷ lệ**, và (3) **thay đổi phần trăm** hay **tốc độ tăng trưởng phần trăm**. Như vậy, $(X_t - X_{t-1})$ biểu thị thay đổi tuyệt đối, $(X_t - X_{t-1})/X_{t-1} = (X_t/X_{t-1} - 1)$ là thay đổi tương đối hay tỷ lệ và $[(X_t - X_{t-1})/X_{t-1}]100$ là thay đổi phần trăm, hay tốc độ tăng trưởng, X_t và X_{t-1} tương ứng là các giá trị hiện tại và quá khứ của biến *X*.

[9] Mô hình hệ số co giãn không đổi cho ta một mức thay đổi tổng doanh thu không đổi đối với một tỷ lệ phần trăm thay đổi giá cả cho trước mà không phụ thuộc vào mức giá tuyệt đối. Người đọc phải so sánh kết quả này với các điều kiện về độ co giãn suy từ hàm hồi quy tuyến tính đơn giản, $Y_i = \beta_1 + \beta_2 X_i + u_i$. Tuy nhiên, một hàm tuyến tính đơn giản cho ta một mức thay đổi số lượng không đổi khi giá cả thay đổi đi một đơn vị. So sánh nó với trường hợp mô hình lôgarít tuyến tính khi giá cả thay đổi đi 1 USD.

[10] Về bản chất của thiên lệch và làm thế nào để giải quyết, xem Arthur S. Goldberger, *Topics in Regression Analysis* (Các chủ đề trong phân tích hồi quy), Macmillan, New York, 1978, trang 120.

Trong mô hình hai biến, cách đơn giản nhất để quyết định xem mô hình tuyến tính lôgarít có thích hợp với số liệu hay không là vẽ lên đồ thị phân tán biểu diễn lnY_i theo lnX_i và xem xem nếu các điểm phân tán nằm gần đúng theo một đường thẳng, như trong Hình 6.3b.



HÌNH 6.3
Mô hình hệ số co giãn không đổi

Ví dụ minh họa: Quay lại với hàm cầu cà phê

Tham chiếu hàm cầu cà phê trong Mục 3.7. Trợ lý nghiên cứu của tôi đã báo cho tôi rằng khi số liệu được vẽ sử dụng tỷ lệ ln$Y$ và ln$X$, đồ thị phân tán chỉ ra rằng mô hình log-log cho ta một sự thích hợp như mô hình tuyến tính (3.7.1).[11] Thực hiện tính toán, người trợ lý thu được các kết quả sau:

$$\begin{array}{rclll}
\ln \hat{Y}_t & = & 0,7774 \quad - & 0,2530 \ln X_t & r^2 = 0,7448 \\
 & & (0,0152) & (0,0494) & F_{1,9} = 26,27 \\
t & = & (51,1447) & (-5,1214) & \\
\textbf{Giá trị p} & = & (0,000) & (0,0003) &
\end{array} \qquad (6.4.5)$$

với Y_t = tiêu dùng cà phê, ly/người/ngày, và X_t = giá thực của cà phê, USD/pao.

Từ các kết quả này, ta thấy hệ số co giãn giá cả là –0,25, có nghĩa là với 1% gia tăng mức giá thực của 1 pao cà phê, mức cầu cà phê (tính bằng số ly cà phê tiêu dùng một ngày) bình quân giảm đi 0,25%. Do giá trị hệ số co giãn giá cả là 0,25 nhỏ hơn 1 về giá trị tuyệt đối, ta có thể nói rằng cầu cà phê không có tính co giãn về giá cả.

Một câu hỏi thú vị: So sánh các kết quả của hàm cầu lôgarít tuyến tính với hàm cầu tuyến tính (3.7.1), làm sao ta quyết định mô hình nào tốt hơn? Ta có thể nói (6.4.5) tốt hơn (3.7.1) bởi vì giá trị r^2 của nó lớn hơn (0,7448 so với 0,6628) được không? Tuy nhiên, ta không thể phát biểu như vậy, bởi vì như sẽ giải thích trong Chương 7, khi biến phụ thuộc của hai mô hình không giống nhau (ở đây là ln$Y$ so với Y), hai giá trị r^2 không so sánh trực tiếp được. Ta cũng không thể so sánh trực tiếp hai hệ số độ dốc, do trong (3.7.1) hệ số độ dốc biểu thị tác động khi giá cà phê thay đổi đi một đơn vị, ví dụ 1USD/pao đối với số lượng giảm sút tuyệt đối không đổi (nghĩa là không phải tương đối) trong tiêu dùng cà phê, trong trường hợp này là 0,4795 ly/ngày. Mặt khác, hệ số độ dốc –0,2530 tính được trong (6.4.5) cho ta tỷ lệ phần

[11] Tất nhiên là (3.7.1) được giới thiệu hoàn toàn với mục đích sư phạm.

trăm giảm sút không đổi trong tiêu dùng cà phê khi giá 1 pao cà phê tăng lên 1% (nghĩa là nó biểu thị độ co giãn giá cả).[12]

Làm sao ta có thể so sánh các kết quả của hai mô hình? Câu hỏi này là một phần của một vấn đề lớn hơn là **phân tích đặc trưng** (specification analysis), một chủ đề sẽ được thảo luận trong Chương 13. Bây giờ, một cách để ta có thể so sánh hai mô hình là tính một đại lượng gần đúng của hệ số co giãn giá cả cho mô hình (3.7.1). Điều đó có thể được thực hiện như sau:

Hệ số co giãn E của biến Y (ví dụ lượng cầu) đối với một biến khác X được định nghĩa là:

$$\begin{aligned} E &= \frac{\%\ \text{thay đổi}\ Y}{\%\ \text{thay đổi}\ X} \\ &= \frac{(\Delta Y / Y)\cdot 100}{(\Delta X / X)\cdot 100} \\ &= \frac{\Delta Y}{\Delta X}\cdot\frac{X}{Y} \\ &= (\text{hệ số độ dốc})(X/Y) \end{aligned} \tag{6.4.6}$$

với Δ biểu thị thay đổi (nhỏ). Nếu Δ đủ nhỏ, ta có thể thay thế $\Delta Y/\Delta X$ bởi dạng đạo hàm, dY/dX (xem chú thích 8).

Bây giờ đối với mô hình tuyến tính (3.7.1), ước lượng của hệ số độ dốc được tính bởi hệ số ước lượng β_2, trong hàm cầu cà phê là –0,4795. Như (6.4.6) biểu thị, để tính độ co giãn, ta phải nhân hệ số độ dốc này với tỷ lệ (X/Y), tức là giá cả chia cho số lượng. Nhưng ta chọn giá trị nào của X và Y? Như Bảng 3.4 biểu thị, có 11 cặp giá trị giá cả (X) và số lượng (Y). Nếu ta sử dụng tất cả các giá trị này, ta sẽ có 11 ước lượng của độ co giãn giá cả.

Tuy nhiên trên thực tế, hệ số co giãn được tính bằng giá trị trung bình hay bình quân của Y và X. Tức là, ta có một ước lượng về hệ số co giãn *trung bình*. Trong ví dụ của chúng ta, $\overline{Y}$ = 2,43 ly và $\overline{X}$ = 1,11 USD. Sử dụng các giá trị này và hệ số độ dốc –0,4795, từ (6.4.6) ta có hệ số co giãn giá cả trung bình là: (–0,4795)(1,11/2,43) = –0,219, hay khoảng –0,22. Kết quả này được so sánh với hệ số co giãn khoảng –0,25 tính từ mô hình tuyến tính lôgarít. Lưu ý rằng hệ số co giãn trong trường hợp sau giữ nguyên không phụ thuộc vào mức giá để tính (tại sao?), trái lại hệ số co giãn trong trường hợp trước phụ thuộc vào các giá trị trung bình cụ thể.

[12] Có sự khác nhau giữa hệ số độ dốc và hệ số co giãn. Như chú thích 8 giải thích, hệ số co giãn bằng hệ số độ dốc (=dY/dX) nhân với tỷ lệ (X/Y). Hệ số độ dốc của mô hình (3.7.1) chỉ là (dY/dX), trái lại hệ số độ dốc trong (6.4.5) cho ta hệ số co giãn (dY/dX)(X/Y). Nói ngắn gọn, **đối với mô hình tuyến tính lôgarít, hệ số độ dốc và hệ số co giãn là như nhau, nhưng khác nhau trong mô hình tuyến tính.**

6.5 CÁC MÔ HÌNH BÁN LOGARIT (SEMILOG): CÁC MÔ HÌNH LOG-LIN VÀ LIN-LOG

Làm thế nào để đo tốc độ tăng trưởng: Mô hình Log-Lin

Các nhà kinh tế, nhà kinh doanh, và chính phủ thường quan tâm tới việc xác định tốc độ tăng trưởng của một số biến kinh tế nhất định, như dân số, GNP, lượng cung tiền, việc làm, năng suất, thâm hụt thương mại, v.v…

Trong bài tập 2.2 ta đã trình bày số liệu GDP của Hoa Kỳ trong giai đoạn 1972-1991. Giả sử ta muốn tìm tốc độ tăng trưởng GDP thực trong giai đoạn này. Đặt Y_t = GDP thực (RGDP) vào thời điểm t và Y_O = giá trị ban đầu (năm 1972) của GDP thực. Bây giờ nhớ lại công thức tính lãi suất gộp nổi tiếng trong các khóa học giới thiệu về tiền tệ, tài chính và ngân hàng.

$$Y_t = Y_O(1 + r)^t \tag{6.5.1}$$

với r là tốc độ tăng trưởng gộp (theo thời gian) của Y. Lấy lôgarít tự nhiên của (6.5.1), ta có thể viết

$$\ln Y_t = \ln Y_O + t\ln(1 + r) \tag{6.5.2}$$

Bây giờ đặt

$$\beta_1 = \ln Y_O \tag{6.5.3}$$

$$\beta_2 = \ln(1 + r) \tag{6.5.4}$$

ta có thể viết (6.5.2) dưới dạng

$$\ln Y_t = \beta_1 + \beta_2 t \tag{6.5.5}$$

Cộng thêm yếu tố nhiễu vào (6.5.5), ta có[13]

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \tag{6.5.6}$$

Mô hình này giống mọi mô hình tuyến tính khác ở chỗ các thông số β_1 và β_2 là tuyến tính. Sự khác nhau duy nhất là biến hồi quy phụ thuộc là lôgarít của Y và biến hồi quy độc lập là "thời gian", lấy giá trị 1, 2, 3, v.v…

Các mô hình như (6.5.6) được gọi là **mô hình bán lôgarít (semilog)** do chỉ có một biến (trong trường hợp này là biến hồi quy phụ thuộc) xuất hiện dưới dạng lôgarít. Đối với các mục đích mô tả, một mô hình trong đó biến hồi quy phụ thuộc được lôgarít hóa sẽ được gọi là **mô hình log-lin**. Chúng ta sẽ xem xét mô hình trong đó biến hồi quy phụ thuộc là tuyến tính nhưng biến hồi quy độc lập được lôgarít hóa ở phần sau và gọi nó là **mô hình lin-log**.

Trước khi ta trình bày các kết quả hồi quy, hãy xem xét các tính chất của mô hình (6.5.5). Trong mô hình này *hệ số độ dốc đo sự thay đổi tỷ lệ hay tương đối không đổi của Y*

[13] Ta đưa thêm sai số vào bởi vì công thức lãi suất gộp sẽ không thoả mãn chính xác. Tại sao ta lại cộng sai số sau khi đã đổi lôgarít sẽ được giải thích trong Mục 6.8.

đối với một sự thay đổi tuyệt đối cho trước về giá trị của biến giải thích (trong trường hợp này là biến *t*), tức là,[14]

$$\beta_2 = \frac{\text{Thay đổi tương đối của biến hồi quy phụ thuộc}}{\text{Thay đổi tuyệt đối của biến hồi quy độc lập}} \qquad (6.5.7)$$

Nếu nhân thay đổi tương đối của *Y* lên 100, (6.5.7) sẽ cho ta thay đổi phần trăm, hay *tốc độ tăng trưởng*, của *Y* đối với thay đổi tuyệt đối của *X*, biến hồi quy độc lập.

Một mô hình log-lin giống như (6.5.5) đặc biệt hữu ích trong các trường hợp mà biến *X* là thời gian, như trong ví dụ GNP của chúng ta, bởi vì trong trường hợp đó, mô hình mô tả tỉ lệ tương đối không đổi (= β_2) hay phần trăm không đổi ($100\cdot\beta_2$) *tốc độ tăng trưởng* (nếu $\beta_2 > 0$) hay *tốc độ giảm sút* (nếu $\beta_2 < 0$). Đó là lý do tại sao các mô hình như (6.5.5) được gọi là **mô hình tăng trưởng (không đổi)**.

Trở lại ví dụ GDP thực, ta có thể viết các kết quả hồi quy dựa vào (6.5.6) như sau:

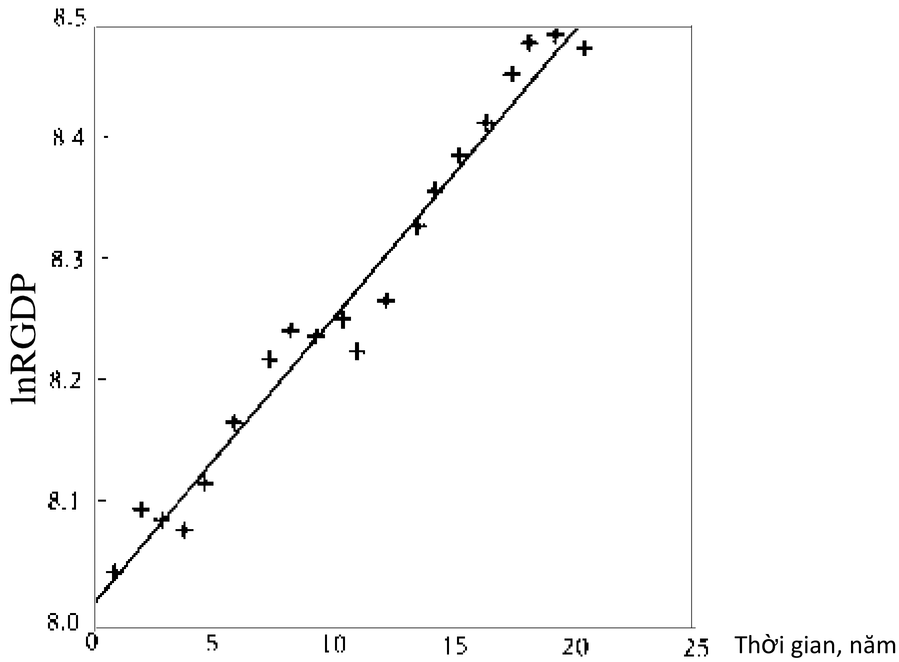
$$\begin{array}{rclcl} \widehat{\ln RGDP}_t & = & 8{,}0139 & + & 0{,}02469t \\ \text{se} & = & (0{,}0114) & & (0{,}00956) \quad r^2 = 0{,}9738 \\ t & = & (7{,}0054) & & (25{,}8643) \\ \text{Giá trị } p & = & (0{,}0000)^* & & (0{,}0000)^* \end{array} \qquad (6.5.8)$$

* biểu thị một giá trị rất nhỏ

Hồi quy này được giải thích như sau: trong giai đoạn 1972-1991, GDP thực tại Hoa Kỳ tăng với tốc độ 2,469%/năm. Từ 8,0139 = $\ln Y_o$ (tại sao?), nếu lấy đối lôgarit của 8,0139, ta tìm được $\hat{Y}_o$ = 3022,7 (gần đúng), tức là vào đầu năm 1972 GDP thực ước lượng vào khoảng 3023 tỷ USD. Đường hồi quy thu được trong (6.5.8) được vẽ trong Hình 6.4.

[14] Sử dụng giải tích vi phân, ta có thể chỉ ra rằng $\beta_2 = d(\ln Y)/dX = (1/Y)(dY/dX) = (dY/Y)/dX$. Công thức này chính là (6.5.7). Đối với các thay đổi nhỏ của *Y* và *X*, quan hệ này được được tính gần đúng bởi:

$$\frac{(Y_t - Y_{t-1})/Y_{t-1}}{(X_t - X_{t-1})}$$



HÌNH 6.4
Tăng trưởng GDP thực, Hoa Kỳ, 1972-1991; mô hình bán lôgarit (semilog).

Tốc độ tăng trưởng tức thời so với tốc độ tăng trưởng gộp. Hệ số độ dốc 0,02469 tính được trong (6.5.8) hay tổng quát hơn, hệ số độ dốc β_2 của mô hình tăng trưởng (6.5.5) cho ta tốc độ tăng trưởng **tức thời** (một điểm trong khoảng thời gian) chứ không phải tốc độ tăng trưởng **gộp** (trong một khoảng thời gian). Nhưng tốc độ tăng trưởng gộp có thể được tính dễ dàng từ (6.5.4): Đơn giản là lấy antilog của 0,02469, trừ đi 1 và nhân hiệu số với 100. Như vậy, trong ví dụ hiện tại, antilog(0,02469) – 1 = 0,024997 hay khoảng 2,499%. Tức là, trong giai đoạn nghiên cứu, *tốc độ tăng trưởng gộp của GDP thực vào khoảng 2,499%/năm*. Tốc độ tăng trưởng này hơi cao hơn tốc độ tăng trưởng tức thời là 2,469%.

Mô hình xu hướng tuyến tính. Thay cho việc ước lượng mô hình (6.5.6), các nhà nghiên cứu đôi khi ước lượng mô hình sau:

$$Y_t = \beta_1 + \beta_2 t + u_t \qquad (6.5.9)$$

Tức là, thay cho việc tính hồi quy của log*Y* theo thời gian, họ tính hồi quy của *Y* theo thời gian. Mô hình như vậy được gọi là **mô hình xu hướng tuyến tính** và biến thời gian *t* được gọi là **biến xu hướng**. Thuật ngữ *xu hướng* có nghĩa là một dịch chuyển đi lên hay đi xuống bền vững trong hành vi của một biến. Nếu hệ số độ dốc trong (6.5.9) dương, *Y* có *xu hướng đi lên*, trái lại nếu hệ số độ dốc âm, *Y có xu hướng đi xuống*.

Các kết quả dựa vào (6.5.9) với số liệu GDP thực của chúng ta được biểu diễn như sau:

$$\begin{aligned} \widehat{\ln RGDP_t} &= 2933{,}0538 + 97{,}6806t \\ \text{se} &= (50{,}5913) \qquad (4{,}2233) \qquad r^2 = 0{,}9674 \\ t &= (57{,}9754) \qquad (231291) \\ \text{Giá trị } p &= (0{,}0000)^* \qquad (0{,}0000)^* \end{aligned} \tag{6.5.10}$$

* biểu thị một giá trị rất nhỏ.

Tương phản với (6.5.8), mô hình hồi quy này được giải thích như sau. Trong giai đoạn 1972 đến 1991, bình quân, GDP thực tăng với một tốc độ tuyệt đối (*Lưu ý*: không phải tương đối) khoảng 97,68 tỷ USD. Như vậy, trong giai đoạn đó, GDP thực có xu hướng đi lên.

Sự lựa chọn giữa mô hình tăng trưởng (6.5.8) và mô hình xu hướng tuyến tính (6.5.10) sẽ phụ thuôc vào việc ta quan tâm tới thay đổi tương đối hay tuyệt đối của GDP thực, mặc dù đối với nhiều mục đích thì thay đổi tương đối quan trọng hơn. Trong khi chuyển sang phần khác, lưu ý rằng *ta không thể so sánh các giá trị r^2 của mô hình (6.5.8) và (6.5.10) bởi vì các biến hồi quy phụ thuộc trong hai mô hình khác nhau*.

Một thận trọng đối với các mô hình log-lin và xu hướng tuyến tính. Mặc dù các mô hình này được sử dụng khá thường xuyên để ước lượng thay đổi tương đối hay tuyệt đối của biến phụ thuộc theo thời gian, việc sử dụng chúng thường xuyên vì mục đích này đã bị các nhà phân tích chuỗi thời gian đặt câu hỏi. Lập luận chủ yếu của họ là những mô hình như vậy chỉ thích hợp nếu chuỗi thời gian có trạng thái tĩnh theo như định nghĩa trong Mục 1.7. Đối với người đọc trình độ cao vấn đề này được thảo luận rất chi tiết trong Chương 21 về Kinh tế lượng Chuỗi thời gian (*lưu ý*: đây là một chương không bắt buộc).

Mô hình lin-log

Giả sử ta có số liệu như trong Bảng 6.3, với *Y* là GNP và *X* là lượng cung tiền (theo định nghĩa M_2). Tiếp theo, giả sử ta quan tâm tới việc tìm xem GNP tăng lên bao nhiêu (về giá trị tuyệt đối) nếu lượng cung tiền tăng lên 1%.

Không giống mô hình tăng trưởng vừa thảo luận trong đó ta quan tâm tới việc tìm xem gia tăng phần trăm của *Y* khi *X* tăng lên 1 đơn vị, bây giờ ta quan tâm tới việc tìm sự thay đổi tuyệt đối của *Y* khi *X* thay đổi đi 1%. Một mô hình phục vụ cho mục tiêu này có thể được viết như sau:

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \tag{6.5.11}$$

Với các mục đích mô tả, ta gọi mô hình như vậy là **mô hình lin-log**.

Bây giờ hãy giải thích hệ số độ dốc β_2.[15] Như thường lệ

[15] Một lần nữa, sử dụng vi phân, ta có $dY/dX = \beta_2(1/X)$. Do đó, $\beta_2 = dY/(dX/X)$ = (6.5.12).

$$\beta_2 = \frac{\text{Thay đổi của } Y}{\text{Thay đổi của } \ln X}$$
$$= \frac{\text{Thay đổi của } Y}{\text{Thay đổi tương đối của } X}$$

Bước thứ hai được suy từ lập luận là *một thay đổi logarít của môt số là thay đổi tương đối*.

Bằng ký hiệu, ta có

$$\beta_2 = \frac{\Delta Y}{\Delta X / X} \tag{6.5.12}$$

với Δ biểu thị một thay đổi nhỏ. Phương trình (6.5.12) có thể được viết một cách tương đương như sau

$$\Delta Y = \beta_2(\Delta X/X) \tag{6.5.13}$$

Phương trình này phát biểu rằng thay đổi tuyệt đối của Y ($=\Delta Y$) bằng β_2 nhân với thay đổi tương đối của X. Nếu thay đổi tương đối của X được nhân với 100, (6.5.13) cho biết thay đổi tuyệt đối của Y đối với thay đổi phần trăm của X. Như vậy, nếu $\Delta X/X$ thay đổi đi 0,01 đơn vị (hay 1%), thay đổi tuyệt đối của Y là 0,01(β_2). Vậy, nếu trong một ứng dụng ta tìm thấy rằng β_2 =500, giá trị tuyệt đối của Y là (0,01)(500), hay 5,0. Do đó, *khi các hồi quy như (6.5.11) được ước lượng bởi OLS, nhân giá trị hệ số độ dốc ước lượng, β_2, với 0,01, hay một cách làm hoàn toàn tương đương là chia cho 100.*

Quay lại với số liệu trong Bảng 6.3, ta có thể viết các kết quả hồi quy của chúng ta như sau:

$$\begin{aligned} \hat{Y}_t &= -16329{,}0 + 2584{,}8X_t \\ t &= (-23{,}494) \quad (27{,}549) \quad r^2 = 0{,}9832 \\ \text{Giá trị } p &= (0{,}0000)^* \quad (0{,}0000)^* \end{aligned} \tag{6.5.14}$$

* biểu thị một giá trị rất nhỏ.

BẢNG 6.3
GNP và lượng cung tiền, Hoa Kỳ, 1973-1987

Năm	**GNP tỷ USD**	**M_2**
1973	1.359,3	861,0
1974	1.472,8	908,5
1975	1.598,4	1023,2
1976	1.782,8	1163,7
1977	1.990,5	1286,7
1978	2.249,7	1389,0
1979	2.508,2	1500,2
1980	2.723,0	1633,1
1981	3.052,6	1795,5
1982	3.166,0	1954,0

1983	3.405,7	2185,2
1984	3.772,2	2363,6
1985	4.014,9	2562,6
1986	4.240,3	2807,7
1987	4.526,7	2901,0

Lưu ý: Các số liệu GNP là số liệu hàng quý trên cơ sở tốc độ hàng năm đã hiệu chỉnh theo mùa.

M_2 = tiền mặt + tiền gửi không kỳ hạn + séc du lịch + các loại tiền gửi được rút séc khác + hợp đồng mua lại chứng khoán (RP) 1 ngày đêm và Eurodollar + số dư MMMF (quỹ hỗ tương trên thị trường tiền tệ) + MMDAs (các tài khoản tiền gửi trên thị trường tiền tệ) + tiết kiệm và tiền gửi nhỏ

Đây là các số liệu trung bình hàng ngày, đã hiệu chỉnh theo mùa.

Nguồn: Báo cáo kinh tế của Tổng thống, 1989, số liệu GNP lấy từ Bảng B-1, trang 308, và số liệu M_2 từ Bảng B-67, trang 385.

Lưu ý rằng ta không đưa ra sai số chuẩn (bạn có thể tính chúng được không?).

Giải thích theo cách vừa trình bày, hệ số độ dốc khoảng 2585 có nghĩa là trong khoảng thời gian của mẫu, lượng cung tiền tăng lên 1%, bình quân, kéo theo sự gia tăng GNP khoảng 25,85 tỷ USD. (*lưu ý*: chia hệ số độ dốc được ước lượng cho 100).

Trước khi tiếp tục, lưu ý rằng nếu muốn tính hệ số co giãn cho các mô hình log-lin hay lin-log, ta có thể thực hiện từ định nghĩa hệ số co giãn ở trên, cụ thể, $(dY/dX)(X/Y)$. Trên thực tế, khi biết dạng hàm số của mô hình, ta có thể tính các hệ số co giãn bằng cách áp dụng định nghĩa ở trên. Bảng 6.5, trình bày ở phần sau, tóm tắt các hệ số co giãn cho các mô hình khác nhau mà ta đã xem xét trong chương này.

6.6 MÔ HÌNH NGHỊCH ĐẢO

Các mô hình có dạng sau được gọi là mô hình **nghịch đảo**.

$$Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i \tag{6.6.1}$$

Mặc dù mô hình này là phi tuyến theo biến X bởi vì biến X có dạng ngược hay nghịch đảo, mô hình có dạng tuyến tính theo β_1 và β_2 và do vậy mô hình là mô hình hồi quy tuyến tính.[16]

Mô hình này có các đặc điểm sau: Khi X tiến dần tới vô cùng, số hạng $\beta_2(1/X)$ dần tới không (*lưu ý*: β_2 không đổi) và Y tiến tới giá trị giới hạn hay tiệm cận β_1. Do vậy, các mô hình như (6.6.1) tạo nên một giá trị **tiệm cận** hay giới hạn mà biến phụ thuộc sẽ nhận khi giá trị của biến X dần tới vô cùng.[17]

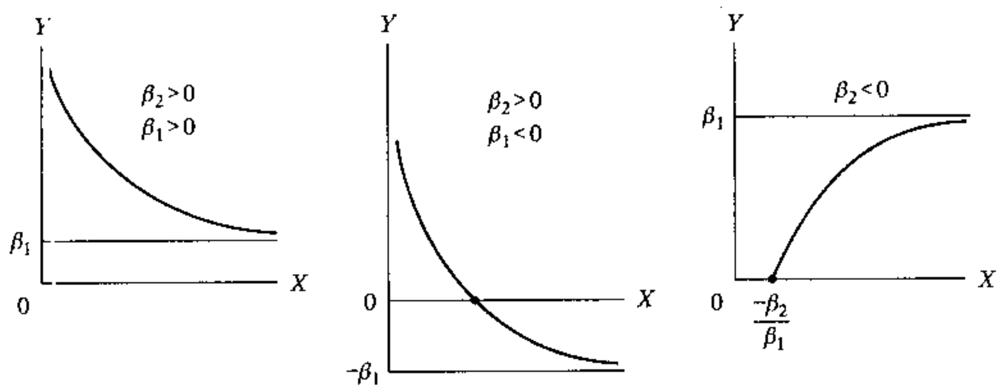
Một số hình dạng thường gặp của đường cong tương ứng với (6.6.1) được biểu diễn trong Hình 6.5. Một ví dụ trong Hình 6.5a được đưa ra trong Hình 6.6, trong đó chi phí sản

[16] Nếu ta gọi $X_i^* = (1/X_i)$, (6.6.1) có các thông số và các biến Y_i và X_i^* tuyến tính.

[17] Độ dốc của (6.6.1) là: $dY/dX = -\beta_2(1/X^2)$, suy ra nếu β_2 dương, độ dốc âm, và nếu β_2 âm độ dốc dương. Xem Hình 6.5a và 6.5c tương ứng.

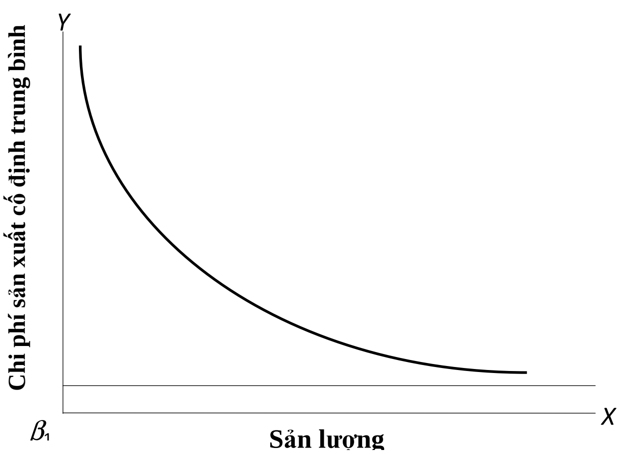
xuất cố định trung bình (AFC) quan hệ với sản lượng. Như hình vẽ biểu thị, AFC giảm liên tục khi sản lượng tăng (do chi phí cố định được chia cho số lượng lớn các đơn vị sản lượng) và cuối cùng sẽ trở nên tiệm cận với trục sản lượng ở mức β_1.

Một trong các ứng dụng quan trọng của Hình 6.5b là đường cong Phillips trong kinh tế vĩ mô. Dựa vào số liệu về tỷ lệ phần trăm thay đổi mức lương (*Y*) và tỷ lệ thất nghiệp tính theo phần trăm (*X*) của Anh Quốc trong giai đoạn 1861- 1957, Phillips thu được một đường cong có dạng tổng quát giống với Hình 6.5b và được tái lập trong Hình 6.7.[18]



HÌNH 6.5

Mô hình nghịch đảo: $Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$.

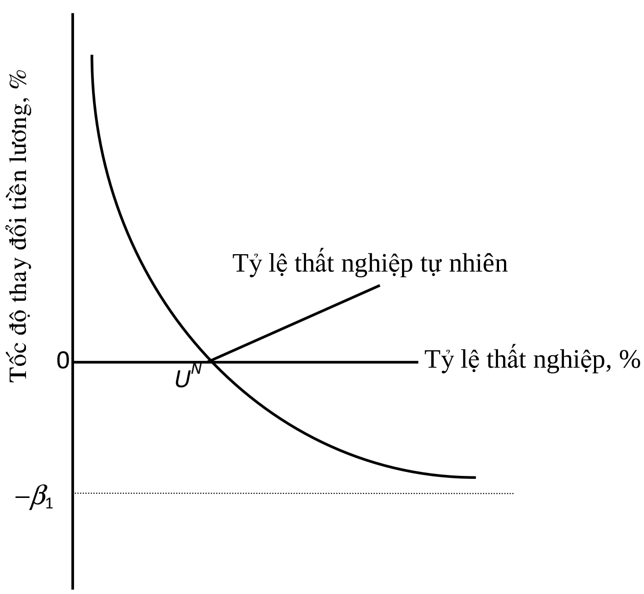


HÌNH 6.6
Mô hình nghịch đảo

[18] A. W. Phillips, "The Relation between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957" (Quan hệ giữa thất nghiệp và tốc độ thay đổi mức lương tại Anh Quốc, 1861-1957), *Economica*, 11/1958, tập 25, trang 283-299. Lưu ý rằng đường cong nguyên thủy được tính thích hợp với số liệu trong giai đoạn 1861 đến 1913 và không cắt trục thất nghiệp, nhưng Hình 6.7 biểu diễn bức tranh hiện đại của lý thuyết Phillips.

Như Hình 6.7 biểu diễn, có sự bất cân xứng trong sự phản ứng của những thay đổi về mức lương đối với tỷ lệ thất nghiệp: lương tăng nhanh khi tỷ lệ thất nghiệp thay đổi một đơn vị nếu tỷ lệ thất nghiệp ở dưới mức U^N, được các nhà kinh tế gọi là *tỷ lệ thất nghiệp tự nhiên* so với khi mức lương giảm xuống đối với một sự thay đổi tương đương của tỷ lệ thất nghiệp khi tỷ lệ này ở trên mức tự nhiên. β_1 biểu thị một giới hạn tiệm cận tương đối với thay đổi của mức lương. Đặc điểm cá biệt này của đường cong Phillips có thể là do các yếu tố định chế như quyền lực thương lượng của công đoàn, mức lương tối thiểu, trợ cấp thất nghiệp, v.v…

Một ứng dụng quan trọng của Hình 6.5c là đường chi tiêu Engel (lấy tên của nhà thống kê người Đức Ernst Engel, 1821-1896). Đường chi tiêu Engel biểu diễn quan hệ giữa chi tiêu của người tiêu dùng cho một hàng hóa với tổng chi tiêu hay thu nhập của người đó. Nếu ta gọi *Y* là chi tiêu cho một loại hàng hóa và *X* là thu nhập thì một số hàng hóa các các đặc điểm sau: (a) Có một mức thu nhập tới hạn hay ngưỡng mà dưới đó thì người tiêu dùng không mua loại hàng hóa này; trong Hình 6.5c mức thu nhập ngưỡng này là $-(\beta_1/\beta_2)$. (b) Có một mức tiêu dùng bão hòa (đã thỏa mãn) mà cao hơn mức đó người tiêu dùng sẽ không chi tiêu thêm nữa cho dù thu nhập có mức cao thế nào đi nữa. Mức này chính là đường tiệm cận β_1 vẽ trong đồ thị. Đối với những hàng hóa này, mô hình nghịch đảo trình bày trong Hình 6.5c là thích hợp nhất.[19]



HÌNH 6.7
Đường cong Phillips

Ví dụ minh họa: Đường cong Phillips của Anh Quốc, 1950-1966

Bảng 6.4 cho ta số liệu về thay đổi phần trăm hàng năm về mức lương, *Y* và tỷ lệ thất nghiệp, *X* của Anh Quốc trong giai đoạn 1950-1966.

---

[19] Về ví dụ cụ thể, xem S. J. Parais & H. S. Houthakker, *The Analysis of Family Budgets* (Phân tích ngân sách gia đình), Cambridge University Press, London, 1971, chương 7.

Việc xây dựng một mô hình nghịch đảo (6.6.1) thích hợp với chuỗi số liệu cho ta các kết quả sau (xem kết quả SAS trong Phụ lục 6A, Mục 6A.3):

$$\hat{Y}_t = -1{,}4282 + 8{,}2743\frac{1}{X_t} \qquad r^2 = 0{,}3849 \tag{6.6.2}$$
$$(2{,}0675) \qquad (2{,}8478) \qquad F_{1,15} = 9{,}39$$

với các số trong ngoặc là các sai số chuẩn ước lượng.

Đường hồi quy ước lượng được biểu diễn trong Hình 6.8. Từ hình này ta thấy rõ rằng giới hạn bên dưới của tốc độ thay đổi mức lương là -1,43, tức là khi X tăng lên vô hạn, tỷ lệ phần trăm giảm sút của mức lương sẽ không lớn hơn 1,43%/năm.

Lưu ý rằng giá trị r^2 ước lượng khá thấp nhưng hệ số độ dốc khác 0 về ý nghĩa thống kê, và có dấu đại số đúng. Quan sát này, như ta sẽ lập luận trong Chương 7, là một lý do giải thích tại sao ta không được nhấn mạnh quá mức giá trị r^2.

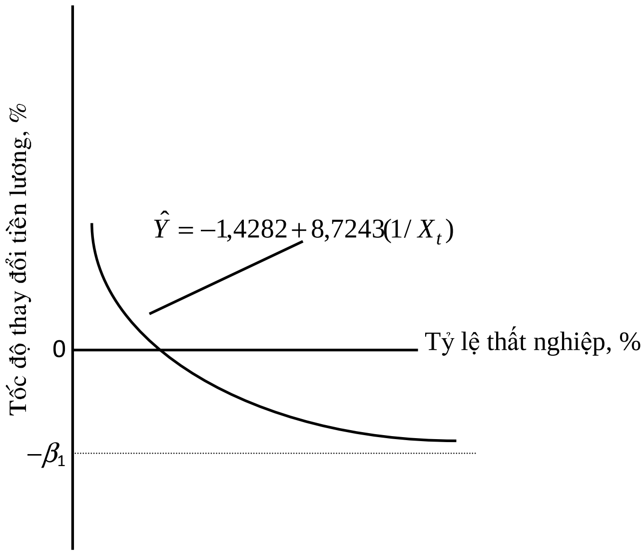
6.7 TÓM TẮT CÁC DẠNG HÀM SỐ

Trong Bảng 6.5 ta tóm tắt các đặc tính nổi bật của các dạng hàm khác nhau mà chúng ta đã phân tích từ đầu cho tới đây.

BẢNG 6.4
Tỷ lệ tăng lương hàng năm và tỷ lệ thất nghiệp, Anh Quốc, 1950-1966

Năm	Tăng lương hàng năm, % *Y*	Thất nghiệp, % *X*
1950	1,8	1,4
1951	8,5	1,1
1952	8,4	1,5
1953	4,5	1,5
1954	4,3	1,2
1955	6,9	1,0
1956	8,0	1,1
1957	5,0	1,3
1958	3,6	1,8
1959	2,6	1,9
1960	2,6	1,5
1961	4,2	1,4
1962	3,6	1,8
1963	3,7	2,1
1964	4,8	1,5
1965	4,3	1,3
1966	4,6	1,4

Nguồn: Cliff Prateen, *Applied Macroeconomics* (Kinh tế vĩ mô ứng dụng), Oxford University Press, Oxford, 1985, trang 85.



HÌNH 6.8
Đường cong Phillips của Anh Quốc, 1950-1966.

BẢNG 6.5

Mô hình	**Phương trình**	**Độ dốc**	**Độ co giãn**
Tuyến tính	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2\left(\frac{X}{Y}\right)$*
Tuyến tính log hay log-log	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{Y}{X}\right)$	β_2
Log-lin	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2(Y)$	$\beta_2(X)$*
Lin-log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{1}{Y}\right)$*
Nghịch đảo	$Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$	$-\beta_2\left(\frac{1}{X^2}\right)$	$-\beta_2\left(\frac{1}{XY}\right)$*

Lưu ý: * có nghĩa là hệ số co giãn thay đổi, phụ thuộc vào giá trị của X hay Y hay cả hai. Khi các giá trị X và Y không được cụ thể hóa, trên thực tế, thường thì các hệ số co giãn này được tính từ giá trị trung bình, $\bar{X}$ và $\bar{Y}$.

* 6.8 MỘT LƯU Ý VỀ BẢN CHẤT CỦA SỐ HẠNG SAI SỐ NGẪU NHIÊN: SỐ HẠNG SAI SỐ NGẪU NHIÊN TỔNG SO VỚI SAI SỐ NGẪU NHIÊN TÍCH

Xem xét mô hình hồi quy sau, giống với (6.4.1) nhưng không có số hạng sai số:

$$Y_i = \beta_1 X^{\beta_2} \tag{6.8.1}$$

Để ước lượng, ta có thể biểu diễn mô hình này dưới ba dạng:

$$Y_i = \beta_1 X_i^{\beta_2} u_I \tag{6.8.2}$$
$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \tag{6.8.3}$$
$$Y_i = \beta_1 X_i^{\beta_2} + u_i \tag{6.8.4}$$

Lấy lôgarít cả hai vế của các phương trình này, ta có

$$\ln Y_i = \alpha + \beta_2 \ln X_i + \ln u_i \tag{6.8.2a}$$
$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \tag{6.8.3a}$$
$$\ln Y_i = \ln(\beta_1 X_i^{\beta_2} + u_i) \tag{6.8.4a}$$

với $\alpha = \ln\beta_1$.

Các mô hình như (6.8.2) là các mô hình về bản chất là tuyến tính (theo thông số) trên khía cạnh là bằng các biến đổi (lôgarít) thích hợp các mô hình này có thể tuyến tính theo các thông số α và β_2. (*Lưu ý*: β_1 trong các mô hình này là phi tuyến). Nhưng mô hình (6.8.4) *về mặt bản chất là phi tuyến theo thông số*. Không có cách đơn giản nào để lấy log của (6.8.4) bởi vì $\ln(A+B) \neq \ln A + \ln B$.

Mặc dù (6.8.2) và (6.8.3) là các mô hình hồi quy tuyến tính và có thể được ước lượng bởi OLS hay ML, ta phải cẩn thận về các tính chất của sai số ngẫu nhiên trong mô hình. Nhớ rằng tính chất ước lượng tuyến tính không thiên lệch tốt nhất (BLUE) của OLS yêu cầu rằng u_i có giá trị trung bình bằng không, phương sai không đổi, và không tự tương quan. Trong kiểm định giả thiết, ta phải giả sử thêm rằng u_i tuân theo phân phối chuẩn với các giá trị trung bình và phương sai vừa thảo luận. Nói ngắn gọn, ta phải giả thiết $u_i \sim N(0,\sigma^2)$.

Bây giờ xem xét mô hình (6.8.2). Đối tác thống kê của nó được trình bày trong (6.8.2a). Để sử dụng mô hình cổ điển về hồi quy tuyến tính chuẩn (CNLRM), ta phải giả thiết rằng

$$\ln u_i \sim N(0,\sigma^2) \tag{6.8.5}$$

Do vậy, khi ta chạy hồi quy (6.8.2a), ta sẽ phải áp dụng các kiểm định quy luật chuẩn thảo luận trong Chương 5 cho các phần dư tính được từ hồi quy này. Nhân tiện, lưu ý rằng nếu $\ln u_i$ tuân theo phân phối chuẩn với giá trị trung bình bằng 0 và phương sai không đổi, thì lý thuyết thống kê cho thấy u_i trong (6.8.2) phải tuân theo **phân phối chuẩn lôgarít** với giá trị trung bình là $e^{\sigma^2/2}$ và phương sai là $e^{\sigma^2}(e^{\sigma^2} - 1)$.

Như phân tích ở trên cho thấy, ta phải rất để ý tới số hạng sai số trong khi biến đổi một mô hình trong phân tích hồi quy. Như đối với (6.8.4), mô hình này là mô hình hồi quy

* Không bắt buộc.

phi tuyến theo thông số và chỉ được giải bằng một chương trình máy tính. Mô hình (6.8.3) sẽ không gây khó khăn gì cho việc ước lượng.

Tóm lại, ta phải rất lưu ý tới số hạng nhiễu khi biến đổi mô hình trong phân tích hồi quy. Nếu không, một sự áp dụng mù quáng phương pháp OLS để biến đổi mô hình sẽ không tạo ra một mô hình với các tính chất thống kê mong muốn.

6.9 TÓM TẮT VÀ KẾT LUẬN

Chương 6 giới thiệu một số đặc điểm sâu hơn của mô hình hồi quy tuyến tính cổ điển (CLRM).

1. Đôi khi một mô hình hồi quy có thể không chứa tung độ gốc. Những mô hình như vậy được gọi là **hồi quy qua gốc tọa độ**. Mặc dù phương pháp đại số để ước lượng các mô hình này đơn giản nhưng ta phải cẩn thận khi sử dụng chúng. Trong những mô hình đó, tổng các phần dư $\sum \hat{u}_i$ khác không; hơn nữa, hệ số r^2 tính theo quy ước có thể không có ý nghĩa. Trừ khi có những lý do lý thuyết mạnh, tốt hơn là đưa tung độ gốc vào trong mô hình một cách rõ ràng.

2. Đơn vị và tỷ lệ mà biến hồi quy phụ thuộc và biến hồi quy độc lập được biểu thị rất quan trọng do sự giải thích các hệ số hồi quy tùy thuộc thiết yếu vào chúng. Trong nghiên cứu thực nghiệm nhà nghiên cứu không nên chỉ có trích dẫn số liệu mà còn phải chỉ rõ các biến được đo lường như thế nào.

3. Dạng hàm số trong quan hệ giữa biến hồi quy phụ thuộc và các biến hồi quy độc lập cũng có ngang tầm quan trọng. Một số dạng hàm số quan trọng thảo luận trong chương này là (a) mô hình tuyến tính logarít hay hệ số co giãn không đổi, (b) mô hình hồi quy semilog, và (c) mô hình nghịch đảo.

4. Trong mô hình tuyến tính lôgarít, cả biến hồi quy phụ thuộc lẫn biến hồi quy độc lập được biểu diện dưới dạng lôgarít. Hệ số hồi quy gắn với log của biến hồi quy độc lập được giải thích là độ co giãn của biến hồi quy phụ thuộc so với biến hồi quy độc lập.

5. Trong mô hình semilog, hoặc biến hồi quy phụ thuộc hoặc biến hồi quy độc lập được biểu diễn dưới dạng lôgarít. Trong mô hình semilog mà biến hồi quy phụ thuộc là lôgarít và biến hồi quy độc lập X là thời gian, hệ số độ dốc ước lượng (nhân với 100) đo tốc độ tăng trưởng (tức thời) của biến hồi quy phụ thuộc. Những mô hình này thường được sử dụng để tính tốc độ tăng trưởng của nhiều hiện tượng kinh tế. Trong mô hình semilog, nếu biến hồi quy độc lập là lôgarít, hệ số của nó đo tốc độ thay đổi tuyệt đối của biến hồi quy phụ thuộc đối với một sự thay đổi phần trăm cho trước của giá trị của biến hồi quy độc lập.

6. Trong các mô hình nghịch đảo, hoặc biến hồi quy phụ thuộc hoặc biến hồi quy độc lập được biểu diễn dưới dạng ngược hay nghịch đảo để tính tới quan hệ phi tuyến giữa các biến số kinh tế, như trong đường cong Phillips.

7. Trong việc lựa chọn các dạng hàm khác nhau, ta phải rất lưu ý tới yếu tố nhiễu ngẫu nhiên u_i. Như đã lưu ý trong Chương 5, mô hình hồi quy tuyến tính cổ điển (CLRM) giả thiết một cách rõ ràng rằng yếu tố nhiễu có giá trị trung bình bằng 0 và phương sai có điều kiện không đổi và rằng nó không tương quan với (các) biến hồi quy độc lập. Với những giả thiết này mà các ước lượng OLS là các ước lượng tuyến tính không thiên lệch tốt nhất (BLUE). Hơn nữa, theo mô hình cổ điển về hồi quy tuyến tính chuẩn, các ước lượng OLS cũng có phân phối chuẩn. Do vậy ta phải tìm xem các giả thiết này có được thỏa mãn không trong dạng hàm số lựa chọn cho phân tích thực nghiệm. Sau khi chạy hồi quy, nhà nghiên cứu phải áp dụng các kiểm định chẩn đoán, như kiểm định quy luật chuẩn, thảo luận trong Chương 5. Điểm này rất quan trọng bởi vì các kiểm định giả thiết cổ điển, như kiểm định t, F, và χ^2, dựa vào giả thiết là yếu tố ngẫu nhiên có phân phối chuẩn. Điều này đặc biệt quan trọng khi kích thước mẫu nhỏ.

8. Mặc dù thảo luận từ trước tới nay được giới hạn ở các mô hình hồi quy hai biến, các chương kế tiếp sẽ chỉ ra rằng trong nhiều trường hợp việc mở rộng ra mô hình hồi quy bội chỉ đơn giản là làm phức tạp hóa phép tính toán chứ không nhất thiết là đưa ra thêm các khái niệm cơ bản. Vì vậy điều quan trọng là người đọc phải nắm thật chắc mô hình hồi quy hai biến.

BÀI TẬP

Câu hỏi

6.1. Xem xét mô hình hồi quy

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

với $y_i = (Y_i - \bar{Y})$ và $x_i = (X_i - \bar{X})$. Trong trường hợp này, đường hồi quy phải đi qua gốc tọa độ. Đúng hay sai? Cho biết cách tính toán của bạn.

6.2. Dựa vào số liệu hàng tháng trong giai đoạn từ 1/1978 đến 12/1987, ta tính được các kết quả hồi quy sau:

$$
\begin{aligned}
\hat{Y}_t &= 0{,}00681 + 0{,}7581X_t \\
\text{se} &= (0{,}02596) \quad (0{,}27009) \\
t &= (0{,}26229) \quad (2.80700) \\
\text{Giá trị } p &= (0{,}7984) \quad (0{,}0186) \qquad r^2 = 0{,}4406
\end{aligned}
$$

$$
\begin{aligned}
\hat{Y}_t &= 0{,}76214X_t \\
\text{se} &= (0{,}265799) \\
t &= (2{,}95408) \\
\text{Giá trị } p &= (0{,}0131) \qquad r^2 = 0{,}43684
\end{aligned}
$$

với Y = suất sinh lợi hàng tháng của cổ phiếu thường của Texaco, %
X = suất sinh lợi hàng tháng của thị trường, %.*

(a) Sự khác nhau giữa hai mô hình hồi quy là gì?
(b) Với các kết quả ở trên, bạn có giữ tung độ gốc trong mô hình đầu tiên không? Tại sao hay tại sao không?

* Số liệu phân tích được lấy từ đĩa số liệu trong Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary* (Thực hành Kinh tế lượng: Cổ điển và Đương đại), Addison-Wesley, Reading, Massachusetts, 1991.

(c) Bạn giải thích hệ số độ dốc trong hai mô hình như thế nào?
(d) Lý thuyết cơ sở hai mô hình là gì?
(e) Bạn có thể so sánh hệ số r^2 trong hai mô hình không? Tại sao có hay tại sao không?
(f) Trị thống kê quy luật chuẩn Jarque-Bera đối với mô hình đầu tiên trong bài tập này là 1,1167 và mô hình thứ hai là 1,1170. Bạn có thể đưa ra kết luận nào từ những trị thống kê này?
(g) Giá trị t của hệ số độ dốc trong mô hình có tung độ gốc bằng 0 vào khoảng 2,95, trong khi trong trường hợp có tung độ gốc là vào khoảng 2,81. Bạn có thể hợp lý hóa kết quả này không?

6.3. Xem xét mô hình hồi quy sau:

$$\frac{1}{Y_i} = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i$$

Lưu ý: Y và X khác 0.
(a) Đây có phải là mô hình hồi quy tuyến tính không?
(b) Làm sao có thể ước lượng được mô hình này?
(c) Hành vi của Y khi X tiến tới vô hạn?
(d) Bạn có thể đưa ra một ví dụ mà mô hình này có thể thích hợp?

6.4. Xem xét mô hình tuyến tính lôgarít:

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

Vẽ Y trên trục tung và X trên trục hoành. Vẽ các đường cong biểu diễn quan hệ giữa Y và X khi $\beta_2 = 1$, $\beta_2 > 1$, và $\beta_2 < 1$.

6.5. **Mô hình lôgarít hypécbôn hay lôgarít nghịch đảo**. Trong bài tập 2.11e ta đã giới thiệu mô hình sau, gọi là mô hình lôgarít nghịch đảo:

$$\ln Y_i = \beta_1 - \beta_2\left(\frac{1}{X_i}\right) + u_i$$

(a) Những tính chất của mô hình này là gì (*Gợi ý*: Xem xét hệ số độ dốc, đường tiệm cận, v.v…)
(b) Gọi X = thời gian. Mô hình này tạo ra loại đường tăng trưởng nào?
(c) Trong những tình thế nào thì bạn sẽ xem xét sử dụng mô hình như thế này?

6.6. Tham chiếu hàm cầu cà phê trong Mục 3.7. Giả sử giá mỗi pao cà phê tính theo xen chứ không phải đô la.
(a) Việc này tác động như thế nào tới tung độ gốc và độ dốc ước lượng trong (3.7.1)? Cho biết các tính toán cần thiết.
(b) Cho biết thay đổi (nếu có) của các sai số chuẩn ước lượng?
(c) r^2 có bị tác động không? Tại sao có hay tại sao không?

6.7. **Hồi quy đối với các biến chuẩn hóa**. Đặt $X_i^* = (X_i - \bar{X}) / S_x$ và $Y_i^* = (Y_i - \bar{Y}) / S_y$, với S_x và S_y là các độ lệch chuẩn tương ứng của X và Y trong mẫu. Hãy chứng tỏ rằng trong mô hình

$$Y_i^* = \alpha_1 + \alpha_2 X_i^* + u_i$$

$\hat{\alpha}_1 = 0$ và $\hat{\alpha}_2 = r$, hệ số tương quan giữa Y và X. Bạn có thể cho biết lý do tại sao người ta lại muốn sử dụng một mô hình hồi quy với các biến chuẩn hóa?

Lưu ý: Y_i^* và X_i^* định nghĩa ở trên được gọi là các **biến chuẩn hóa**. Một biến được gọi là chuẩn hóa hay theo các đơn vị (độ lệch) chuẩn nếu nó được được biểu diễn bằng độ lệch khỏi giá trị trung bình của nó (nghĩa là thay đổi gốc tọa độ) và chia cho độ lệch chuẩn mẫu của nó (nghĩa là thay đổi tỷ lệ). Như vậy, việc chuẩn hóa bao gồm cả thay đổi gốc tọa độ lẫn thay đổi tỷ lệ.

Các biến chuẩn hóa như vậy có các tính chất sau: mỗi biến có giá trị trung bình bằng không và phương sai bằng 1. Kết quả là sự thay đổi đi một đơn vị của X^* trở thành sự thay đổi đi 1 độ lệch chuẩn. Do vậy, hệ số độ dốc trong mô hình hiện tại có thể được giải thích là cho biết số các độ lệch chuẩn mà biến phụ thuôc thay đổi, tính một cách trung bình, khi biến giải thích thay đổi đi 1 độ lệch chuẩn. Một cách ngẫu nhiên, hệ số độ dốc trong mô hình này được là hệ số bêta, nhưng đừng lầm lẫn nó với hệ số bêta trong lý thuyết cơ cấu đầu tư chứng khoán.

6.8. Xem xét các mô hình sau:

Mô hình I: $Y_i = \beta_1 + \beta_2 X_i + u_i$
Mô hình II: $Y_i^* = \alpha_1 + \alpha_2 X_i^* + u_i$

với Y^* và X^* là các biến chuẩn hóa như định nghĩa trong bài tập 6.7. Hãy chỉ ra rằng $\hat{\alpha}_2 = \hat{\beta}_2 (S_x/S_y)$ và từ đó chứng minh rằng *mặc dù các hệ số độ dốc hồi quy độc lập với thay đổi gốc tọa độ chúng không độc lập với thay đổi tỷ lệ*.

6.9. Xem xét các mô hình sau:

$$\ln Y_i^* = \alpha_1 + \alpha_2 \ln X_i^* + u_i^*$$
$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

với $Y_i^* = w_1 Y_i$ và $X_i^* = w_2 X_i$, các w là hằng số.
(a) Thiết lập các quan hệ giữa hai tập hợp các hệ số hồi quy và sai số chuẩn của chúng.
(b) Hệ số r^2 trong hai mô hình có khác nhau không?

6.10. Giữa hồi quy (6.5.8) và (6.5.10), bạn thích chọn mô hình nào? Tại sao?

6.11. Với hàm cầu cà phê ước lượng (6.4.5), bạn có chấp nhận giả thiết cho rằng hệ số co giãn giá cả của mức cầu cà phê không khác 0 về ý nghĩa? Sử dụng kiểm định một phía tại mức ý nghĩa 5%. Xem xét tại sao kiểm định một phía lại thích hợp.

6.12. Đối với hồi quy (6.5.8), kiểm định giả thiết cho rằng hệ số độ dốc không khác 0,03 về ý nghĩa.

6.13. Từ đường cong Phillips ước lượng trong (6.6.2), có thể ước lượng tỷ lệ thất nghiệp tự nhiên được không? Bằng cách nào?

6.14. Đối với mô hình log-lin (6.4.3) nếu Y là số lượng một hàng hóa tiêu dùng và X là thu nhập của người tiêu dùng, làm sao bạn tính được độ co giãn thu nhập: $(dY/dX)(X/Y)$? Và đối với mô hình lin-log (6.5.11)?

6.15. Đường chi tiêu Engel biểu diễn quan hệ giữa chi tiêu của người tiêu dùng đối với một mặt hàng với tổng thu nhập của người đó. Đặt Y = chi tiêu tiêu dùng của một mặt hàng và X = thu nhập của người tiêu dùng, xem xét các mô hình sau:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$Y_i = \beta_1 + \beta_2(1/X_i) + u_i$$
$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_i + u_i$$
$$\ln Y_i = \ln\beta_1 + \beta_2(1/X_i) + u_i$$
$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

Bạn sẽ chọn (những) mô hình nào trong số này cho đường chi tiêu Engel và tại sao? (*Gợi ý*: Giải thích các hệ số độ dốc khác nhau, tìm các cách biểu diễn hệ số co giãn chi tiêu so với thu nhập, v.v…)

Bài tập

6.16. Bảng sau cho biết số liệu của Singapore về chỉ số giảm phát GDP (tổng sản phẩm quốc dân) đối với hàng hóa trong nước và chỉ số giảm phát GDP đối với hàng nhập khẩu trong giai đoạn 1968-1982. Chỉ số giảm phát GDP thường được sử dụng làm một chỉ số cho lạm phát thay cho chỉ số giá hàng tiêu dùng (CPI). Singapore là một nền kinh tế mở quy mô nhỏ, phụ thuộc nhiều vào ngoại thương.

Năm	Chỉ số giảm phát GDP đối với hàng nội địa, *Y*	Chỉ số giảm phát GDP đối với hàng nhập khẩu, *X*
1968	1000	1000
1969	1023	1042
1970	1040	1092
1971	1087	1105
1972	1146	1110
1973	1285	1257
1974	1485	1749
1975	1521	1770
1976	1543	1889
1977	1567	1974
1978	1592	2015
1979	1714	2260
1980	1841	2621
1981	1959	2777
1982	2033	2735

Nguồn: Colin Simkin, “Does Money Matter in Singapore?” (Tiền có là vấn đề gì không ở Singapore), *The Singapore Economic Review* (Tạp chí Kinh tế Singapore), tập XXIX, số 1, 4/1984, Bảng 6, trang 8.

Để nghiên cứu mối quan hệ giữa giá nội địa và thế giới, bạn được cho biết các mô hình sau:

1. $Y_i = \alpha_1 + \alpha_2 X_t + u_i$
2. $Y_t = \beta_2 X_t + u_t$

với Y = chỉ số giảm phát GDP đối với hàng nội địa
X = chỉ số giảm phát GDP đối với hàng nhập khẩu.

(a) Làm thế nào lựa chọn một nghiên cứu trên nghiệm giữa hai mô hình?

(b) Làm thích hợp (ước lượng) cả hai mô hình với số liệu và quyết định xem mô hình nào thích hợp hơn?
(c) Các mô hình nào khác có thể phù hợp với số liệu?

6.17. Tham chiếu số liệu trong Bài tập 6.16. Giá trị trung bình tương ứng của *Y* và *X* là 1456 và 1760, và độ lệch chuẩn tương ứng là 346 và 641. Ước lượng mô hình sau:

$$Y_t^* = \alpha_1 + \alpha_2 X_t^* + u_t$$

với các biến có dấu sao là các biến chuẩn hóa, và giải thích các kết quả.

6.18. Tham chiếu số liệu trong bài tập 1.1. Đối với từng nước, ước lượng tốc độ gia tăng lạm phát từ mô hình:

$$\ln Y_t = \beta_1 + \beta_2 \text{Time} + u_i$$

với *Y* là tỷ lệ lạm phát. Làm sao kết quả này lại khác với các kết quả tính được trong bài tập 5.12?

6.19. Tham chiếu số liệu trong bài tập 3.22. Tính tốc độ tăng trưởng GDP danh nghĩa của Hoa Kỳ trong giai đoạn 1972-1991 và so sánh kết quả của bạn với các kết quả trong Phương trình (6.5.8). Hai kết quả hồi quy này có giúp bạn trong việc ước lượng tỷ lệ lạm phát của Hoa Kỳ trong cùng giai đoạn không? Như thế nào?

6.20. Giả sử bạn làm thích hợp (ước lượng) dạng đường cong Phillips sau với số liệu trong Bảng 6.4:

$$Y_t = \alpha_1 + \alpha_2 X_t + u_t$$

với *Y* = tỷ lệ thay đổi % hàng năm của mức lương
X = tỷ lệ thất nghiệp

(a) Tiên nghiệm dấu của α_2?
(b) Ước lượng hồi quy này, tính các trị thống kê thông thường.
(c) Làm sao các kết quả này so sánh với các kết quả trong hồi quy (6.6.2)? Có mâu thuẫn gì không trong các kết quả?
(d) Có thể so sánh hai giá trị r^2 hay không?
(e) Bạn ưa thích mô hình nào? Tại sao?

6.21. Bảng sau cho biết số liệu GDP và lượng cung tiền (M_1) theo triệu USD của Canada trong giai đoạn 1970-1984.

Năm	**GNP**	**Lượng cung tiền, M_1**
1970	85.685	9.077
1971	94.450	10.178
1972	105.234	11.626
1973	123.560	13.320
1974	147.528	14.555
1975	165.343	16.566
1976	191.857	17.889
1977	210.189	19.381
1978	232.211	21.328
1979	264.279	22.823
1980	297.556	24.254
1981	339.793	25.379
1982	358.302	25.541

1983	390.340	28.137
1984	420.819	28.798

Nguồn: Ngân hàng Dư trữ Liên bang St. Louis, *International Economic Conditions* (Các điều kiện kinh tế quốc tế), Xuất bản hàng năm, 6/1985, trang 14 (số liệu M_1) và trang 17 (số liệu GNP).

Sử dụng những số liệu này để làm thích hợp mô hình sau và bình luận các kết quả.

$$GNP_t = \beta_1 + \beta_2 \ln \text{Tiền}_t + u_t$$

6.22. Bạn được cung cấp số liệu sau:

Y_i	X_i	Y_i	X_i
86	3	62	35
79	7	52	45
76	12	51	55
69	17	51	70
65	25	48	120

Nguồn: Chỉnh sửa từ J. Johnston, *Econometric Methods* (Các Phương pháp Kinh tế Lượng), xuất bản lần thứ 3, McGraw-Hill, New York, 1984, trang 87. Trên thực tế nó được lấy từ một bài thi kinh tế lượng của Đại học Oxford, 1975.

Làm thích hợp mô hình sau với số liệu trong bảng trên và tính các trị thống kê hồi quy thông thường:

$$\left(\frac{100}{100 - Y_i}\right) = \beta_1 + \beta_2\left(\frac{1}{X_i}\right)$$

6.23. Để tính hệ số co giãn thay thế giữa nhập lượng vốn và lao động, Arrow, Chenery, Minhas, và Solow, các tác giả của hàm sản xuất CES (hệ số co giãn thay thế không đổi) nổi tiếng, đã sử dụng mô hình sau:*

$$\log\left(\frac{V}{L}\right) = \log\beta_1 + \beta_2 \log W + u$$

với (V/L) = giá trị gia tăng trên một đơn vị lao động
L = nhập lượng lao động
W = mức lương thực

Hệ số β_2 tính hệ số co giãn thay thế giữa lao động và vốn (nghĩa là thay đổi tỷ lệ của tỷ lệ nhân tố sản xuất/thay đổi tỷ lệ của giá nhân tố sản xuất tương đối). Từ số liệu trong bảng sau, chứng minh rằng hệ số ước lượng là 1,3338 và rằng nó không khác 1 về ý nghĩa thống kê.

* "Capital-Labor Substitution and Econojmic Efficiency", *Review of Economics and Satistics* (Thay thế vốn - lao động và hiệu quả kinh tế, Tạp chí Kinh tế và Thống kê), 8/1961, tập 43, số 5, trang 225-254.

Ngành công nghiệp	**log(*V*/*L*)**	**log*W***
Bột mì	3,6973	2,9617
Đường	3,4795	2,8532
Sơn và vécni	4,0004	3,1158
Xi măng	3,6609	3,0371
Kính và đồng bằng kính	3,2321	2,8727
Gốm	3,3418	2,9745
Gỗ dán	3,4308	2,8287
Dệt bông	3,3158	3,0888
Dệt len	3,5062	3,0086
Dệt đay	3,2352	2,9680
Hóa chất	3,8823	3,0909
Nhôm	3,7309	3,0881
Sắt thép	3,7716	3,2256
Xe đạp	3,6601	3,1025
Máy may	3,7554	3,1354

Nguồn: Damodar Gujarati, "A Test of ACMS Production Function: Indian Industries, 1958", *Indian Journal of Insustrial Relations* (Một kiểm định về Hàm sản xuất ACMS: các ngành công nghiệp Ấn độ, 1958, Tạp chí Quan hệ Lao động Ấn Độ), tập 2, số 1, 7/1996, trang 95-97.

PHỤ LỤC 6A

6A.1 TÍNH CÁC ƯỚC LƯỢNG BÌNH PHƯƠNG TỐI THIỂU CHO HỒI QUY QUA GỐC TỌA ĐỘ

Ta muốn cực tiểu hóa

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_2 X_i)^2 \tag{1}$$

theo $\hat{\beta}_2$.

Lấy vi phân (1) theo $\hat{\beta}_2$, ta có

$$\frac{d\sum \hat{u}_i^2}{d\hat{\beta}_2} = 2\sum (Y_i - \hat{\beta}_2 X_i)(-X_i) \tag{2}$$

Đặt (2) bằng 0 và đơn giản hóa, ta có

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad (6.1.6) = (3)$$

Bây giờ thay thế hàm hồi quy tổng thể (PRF): $Y_i = \beta_2 X_i + u_i$ vào phương trình này, ta có

$$\hat{\beta}_2 = \frac{\sum X_i(\beta_2 X_i + u_i)}{\sum X_i^2} \tag{4}$$

$$= \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2}$$

(*Lưu ý*: $E(\hat{\beta}_2) = \beta_2$). Do đó,

$$E(\hat{\beta}_2 - \beta_2)^2 = E\left[\frac{\sum X_i u_i}{\sum X_i^2}\right]^2 \tag{5}$$

Mở rộng vế phải của (5) và lưu ý rằng X_i không ngẫu nhiên và u_i có phương sai có điều kiện không đổi và không tương quan, ta có

$$\text{var}(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2 = \frac{\sigma^2}{\sum X_i^2} \qquad (6.1.7) = (6)$$

Đồng thời, lưu ý rằng từ (2) sau khi cho nó bằng 0 ta có:

$$\sum \hat{u}_i X_i = 0 \tag{7}$$

Từ Phụ lục 3A, Mục 3A.1 ta thấy rằng khi có tung độ gốc trong mô hình, ta có thêm điều kiện $\sum \hat{u}_i = 0$ vào (7). Từ biểu thức toán học vừa đưa ra ta thấy rõ ràng rằng hồi quy qua gốc tọa độ có thể không có tổng sai số, $\sum \hat{u}_i$, bằng 0.

Giả sử ta muốn đưa ra điều kiện rằng $\sum \hat{u}_i = 0$. Trong trường hợp đó, ta có

$$\sum Y_i = \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \tag{8}$$

$$= \hat{\beta}_2 \sum X_i, \qquad \text{do } \sum \hat{u}_i = 0 \text{ như điều kiện đã đưa ra.}$$

Biểu thức này cho ta

$$\hat{\beta}_2 = \frac{\sum Y_i}{\sum X_i} \tag{9}$$

$$= \frac{\bar{Y}}{\bar{X}} = \frac{\text{giá trị trung bình của } Y}{\text{giá trị trung bình của } X}$$

Nhưng ước lượng này không giống như (3) ở trên hay (6.1.6). Và do $\hat{\beta}_2$ của (3) không thiên lệch (tại sao?), $\hat{\beta}_2$ của (9) không thể không thiên lệch.

Kết quả là, trong hồi quy qua gốc tọa độ, ta không thể có cả $\sum \hat{u}_i X_i$ và $\sum \hat{u}_i$ bằng 0, như trong mô hình quy ước. Điều kiện duy nhất được thỏa mãn là $\sum \hat{u}_i X_i$ bằng 0.

Nhớ lại rằng

$$Y_i = \hat{Y}_i + \hat{u} \tag{2.6.3}$$

Cộng phương trình này ở cả hai vế và chia cho N, cỡ mẫu, ta có

$$\bar{Y} = \bar{\hat{Y}} + \bar{\hat{u}} \tag{10}$$

Do trong mô hình có tung độ gốc bằng 0, $\sum \hat{u}_i$ và do vậy $\bar{\hat{u}}$ không cần phải bằng 0. Từ đó suy ra

$$\bar{Y} \neq \bar{\hat{Y}} \tag{11}$$

tức là, giá trị trung bình của các giá trị Y thực tế không cần phải bằng với giá trị trung bình của các giá trị Y ước lượng; hai giá trị trung bình đồng nhất trong mô hình có tung độ gốc, như có thể thấy trong (3.1.10).

Ta đã lưu ý rằng trong mô hình có tung độ gốc bằng 0, r^2 có thể âm, trái lại trong mô hình quy ước nó luôn luôn không âm. Điều kiện này có thể được giải thích như sau.

Sử dụng (3.5.5a), ta có thể viết

$$r^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \tag{12}$$

Bây giờ đối với mô hình quy ước, hay mô hình có tung độ gốc, Phương trình (3.3.6) cho thấy

$$\text{RSS} = \sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \le \sum y_i^2 \qquad (13)$$

trừ khi $\hat{\beta}_2$ bằng 0 (nghĩa là *X* không có tác động nào tới *Y*). Tức là, đối với mô hình quy ước, RSS ≤ TSS, hay, r^2 không bao giờ âm.

Đối với mô hình có tung độ gốc bằng không ta có thể chỉ ra một cách tương tự là

$$\text{RSS} = \sum \hat{u}_i^2 = \sum Y_i^2 - \hat{\beta}_2^2 \sum X_i^2 \qquad (14)$$

(*Lưu ý*: Các tổng bình phương của *Y* và *X* không hiệu chỉnh theo giá trị trung bình). Bây giờ thì không có đảm bảo rằng RRS sẽ luôn luôn nhỏ hơn $\sum y_i^2 = \sum Y_i^2 - N\bar{Y}^2$ (TSS). Từ đó, RSS có thể lớn hơn TSS, suy ra r^2, như theo định nghĩa quy ước, có thể âm. Đồng thời, lưu ý rằng trong trường hợp này, RSS sẽ lớn hơn TSS nếu $\hat{\beta}_2^2 \sum X_i^2 < N\bar{Y}^2$.

PHỤ LỤC 6A.2 KẾT QUẢ SAS CỦA ĐƯỜNG ĐẶC TÍNH (6.1.12)

BIẾN PHỤ THUỘC: Y

NGUỒN	BẬC TỰ DO	TỔNG CÁC BÌNH PHƯƠNG	BÌNH PHƯƠNG TRUNG BÌNH	GIÁ TRỊ F	XÁC SUẤT > F
MÔ HÌNH	1	12364,263	12364,263	32,375	0,0008
SAI SỐ	9	3437,147	381,905		
TỔNG	10	15801,410			

CĂN MSE	19,542396	R BÌNH PHƯƠNG	0,7825
TRUNG BÌNH DEP	19,390000	R BÌNH PHƯƠNG HIỆU CHỈNH	0,7825
C.V	100,786		

LƯU Ý: KHÔNG SỬ DỤNG SỐ HẠNG TUNG ĐỘ GỐC, R BÌNH PHƯƠNG ĐƯỢC ĐỊNH NGHĨA LẠI

BIẾN	BẬC TỰ DO	ƯỚC LƯỢNG THÔNG SỐ	SAI SỐ CHUẨN	T ĐỐI VỚI H_0: THÔNG SỐ = 0	XÁC SUẤT > $\lvert T \rvert$
X	1	1,089912	0,191551	5,690	0,0008

Số quan sát	Y	X	YHAT	YRESID
1	67,5	19,5	21,253	46,247
2	19,2	8,5	9,264	9,936
3	-35,2	-29,3	-31,394	-3,266
4	-42,0	-26,5	-28,883	-13,117
5	63,7	61,9	67,466	-3,766
6	19,3	45,5	49,591	-30,291
7	3,6	9,5	10,354	-6,754
8	20,0	14,0	15,259	4,741
9	40,3	35,3	38,474	1,826
10	37,5	31,0	33,787	3,713
DURBIN-WATSON			d	0,896
TỰ TƯƠNG QUAN BẬC 1				0,239

Ghi chú: Giá trị xác suất > $\lvert T \rvert$ = 0,0008 đại diện cho giá trị *p*.

PHỤ LỤC 6A.3 KẾT QUẢ SAS CỦA HỒI QUY PHILLIPS TRONG TRƯỜNG HỢP ANH QUỐC (6.6.2)

BIẾN PHỤ THUỘC: Y

NGUỒN	BẬC TỰ DO	TỔNG CÁC BÌNH PHƯƠNG	BÌNH PHƯƠNG TRUNG BÌNH	GIÁ TRỊ F	XÁC SUẤT > F
MÔ HÌNH	1	25,054647	25,054647	9,385	0,0079
SAI SỐ	15	40,043000	2,669533		
TỔNG	16	65,097647			

CĂN MSE	1,633871	R BÌNH PHƯƠNG	0,3849
TRUNG BÌNH DEP	4,788235	R BÌNH PHƯƠNG HIỆU CHỈNH	0,3439
C.V	34,12261		

BIẾN	BẬC TỰ DO	ƯỚC LƯỢNG THÔNG SỐ	SAI SỐ CHUẨN	T ĐỐI VỚI H_0: THÔNG SỐ = 0	XÁC SUẤT > $\lvert T \rvert$
TUNG ĐỘ GỐC	1	-4,8177	2,067478	-0,691	0,5003
X	1	8,724344	2,847779	3,064	0,0079

Số quan sát	Y	X	$X_1 = 1/X$	YHAT	YRESID
1	1,8	1,4	0,71429	4,80350	-3,0035
2	8,5	1,1	0,90909	6,50304	1,9970
3	8,4	1,5	0,66667	4,38805	4,0119
4	4,5	1,5	0,66667	4,38805	0,1119
5	4,8	1,2	0,83333	5,84211	-1,5421
6	6,9	1,0	1,00000	7,29617	-0,3962
7	8,0	1,1	0,90909	6,50304	1,4970
8	5,0	1,3	0,76923	5,28286	-0,2829
9	3,6	1,8	0,055556	3,41868	0,1813
10	2,6	1,9	0,52632	3,16358	-0,5636
11	2,6	1,5	0,66667	4,38805	-1,7881
12	4,2	1,4	0,71429	4,80350	-0,6035
13	3,6	1,8	0,55556	3,41868	0,1813
14	3,7	2,1	0,47619	2,72627	0,9737
15	4,8	1,5	0,66667	4,38805	0,4119
16	4,3	1,3	0,76923	5,28286	-0,9829
17	4,6	1,4	0,71429	4,80350	-0,2035

Ghi chú: Các số dưới tiêu đề xác suất > $\lvert T \rvert$ đại diện cho giá trị *p*.

Chương 7
PHÂN TÍCH HỒI QUY BỘI: VẤN ĐỀ VỀ ƯỚC LƯỢNG

7.10 VÍ DỤ 7.10: HÀM SẢN XUẤT COBB-DOUGLAS : NÓI THÊM VỀ DẠNG HÀM SỐ

Trong Phần 6.4 chúng ta đã cho thấy bằng những phép biến đổi thích hợp chúng ta có thể biến đổi các quan hệ phi tuyến tính thành các quan hệ tuyến tính để chúng ta có thể hoạt động trong khuôn khổ các mô hình hồi qui tuyến tính cổ điển. Các phép biến đổi khác nhau đã được thảo luận đến trong bối cảnh trường hợp hai biến có thể dễ dàng được mở rộng qua cho trường hợp các mô hình hồi qui bội. Trong phần này, chúng ta chứng minh các phép biến đổi này bằng cách mở rộng đa biến của mô hình tuyến tính-logarít hai biến; những trường hợp khác có thể được tìm thấy trong các bài tập và qua các ví dụ minh họa được thảo luận đến trong suốt phần còn lại của tập sách này. Ví dụ cụ thể mà chúng ta thảo luận là **hàm sản xuất Cobb-Douglas** nổi tiếng trong lý thuyết về sản xuất.

Hàm sản xuất Cobb-Douglas , ở dạng ngẫu nhiên của nó, có thể được biểu diễn như sau:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{ui} \tag{7.10.1}$$

trong đó:

Y = sản lượng
X_2 = nhập lượng lao động
X_3 = nhập lượng vốn
u = số hạng nhiễu ngẫu nhiên
e = cơ số của logarít tự nhiên

Từ phương trình (7.10.1) rõ ràng là quan hệ giữa sản lượng và hai yếu tố nhập lượng là không tuyến tính. Tuy nhiên, nếu chúng ta biến đổi logarít mô hình này, chúng ta được

$$\begin{aligned} \ln Y_i &= \ln \beta_1 + \beta_2 \ln X_{3i} + u_i \\ &= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \end{aligned} \tag{7.10.2}$$

trong đó $\beta_0 = \ln\beta_1$.

Viết như vậy, mô hình là tuyến tính theo thông số β_0, β_2 và β_3 và do đó là một mô hình hồi qui tuyến tính. Tuy nhiên lưu ý rằng nó không tuyến tính theo các biến Y và X nhưng tuyến tính theo logarít của các biến này. Tóm lại, (7.10.2) là một *logarít-logarít, logarit kép*, hay mô hình *tuyến tính-logarít*, là mô hình hồi qui bội tương ứng với mô hình tuyến tính logarít hai biến (6.4.3).

Các đặc tính của hàm sản xuất Cobb-Douglas được biết rất rộng rãi:

1. β_2 là độ co dãn (riêng phần) của sản lượng so với nhập lượng lao động, có nghĩa là, nó đo thay đổi phần trăm trong sản lượng ứng với, ví dụ như, thay đổi 1 phần trăm trong nhập lượng lao động, giữ cho nhập lượng vốn không đổi (xem bài tập 7.10).

2. Tương tự như vậy, β_3 là độ co dãn (riêng phần) của sản lượng so với nhập lượng vốn, giữ cho nhập lượng lao động không đổi.

3. Tổng của $(\beta_2 + \beta_3)$ cho ta thông tin về *sinh lơi theo quy mô*, có nghĩa là, sự đáp ứng của sản lượng trước một tỉ lệ thay đổi của nhập lượng. Nếu tổng này bằng 1, thì không có *sinh lợi cố định theo quy mô*, có nghĩa là, tăng gấp hai lần nhập lượng sẽ làm tăng sản lượng lên gấp hai lần, tăng gấp ba lần nhập lượng sẽ làm tăng sản lượng lên gấp ba lần, và cứ vậy. Nếu tổng này nhỏ hơn 1, tức có hiện tượng *sinh lơi giảm dần theo quy mô* -tăng gấp hai lần nhập lượng sẽ không làm sản lượng tăng lên gấp hai. Cuối cùng, nếu tổng này lớn hơn 1, có hiện tượng *sinh lơi tăng dần theo quy mô* - tăng gấp hai lần nhập lượng sẽ làm sản lượng tăng lên hơn gấp hai.

Trước khi tiếp tục, lưu ý rằng bất cứ khi nào các bạn có mô hình hồi qui tuyến tính-logarit với một số lượng biến bất kỳ, hệ số của mỗi biến X là số đo độ co dãn (riêng phần) của biến phụ thuộc Y so với biến X đó. Như vậy, nếu bạn có một mô hình tuyến tính-logarit với k-biến:

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \, . \, . \, . \, . \, . + \beta_k \ln X_{ki} + u_i \qquad (7.10.3)$$

mỗi hệ số hồi qui (riêng phần), β_2 cho đến β_k, là độ co dãn (riêng phần) của Y so với các biến X_2 cho đến X_k.[20]

Để minh họa hàm sản xuất Cobb-Douglas, chúng ta thu thập các dữ liệu trình bày trong Bảng 7.3; những dữ liệu này là của khu vực nông nghiệp của Đài Loan trong giai đoạn 1958-1972.

BẢNG 7.3
Tổng sản lượng thực, ngày lao động và nhập lượng vốn trong khu vực nông nghiệp của Đài Loan, 1958-1972

Năm	Tổng sản lượng thực (triệu NT $)*, Y	Ngày lao động (triệu ngày), X_2	Nhập lượng vốn thực (triệu NT $), X_3
1958	16,607.7	275.5	17,803.7
1959	17,511.3	274.4	18,096.8

[20] Để xem xét điều này, lấy vi phân riêng phần (7.10.3) theo logarít của mỗi biến X. Vì vậy, $\partial \ln Y / \partial \ln X_2 = (\partial Y / \partial X_2)(X_2 / Y) = \beta_2$, giá trị này theo định nghĩa, là độ co dãn của Y theo X_2, và $\partial \ln Y / \partial \ln X_3 = (\partial Y / \partial X_3)(X_3 / Y) = \beta_3$, theo định nghĩa là độ co dãn của Y theo X_3, và v.v.

1960	20,171.2	269.7	18,271.8
1961	20,932.9	267.0	19,167.3
1962	20,406.0	267.8	19,647.6
1963	20,831.6	275.0	20,803.5
1964	24,806.3	283.0	22,076.6
1965	26,465.8	300.7	23,445.2
1966	27,403.0	307.5	24,939.0
1967	28,628.7	303.7	26,713.7
1968	29,904.5	304.7	29,957.8
1969	27,508.2	298.6	31,585.9
1970	29,035.5	295.5	33,474.5
1971	29,281.5	299.0	34,821.8
1972	31,535.8	288.1	41,794.3

Nguồn: Thomas Pei-Fan Chen. "Economic growth and Structural change in Taiwan—1952-1972, A Production Function Approach," luận văn tiến sĩ không xuất bản, Khoa Kinh tế, Graduate Center; City University of New York, June 1976, Bảng II.
* Đôla mới của Đài Loan

Giả định rằng mô hình (7.10.2) thỏa mãn được các giả định của mô hình hồi qui tuyến tính cổ điển,[21] chúng ta lấy hồi qui sau bằng phương pháp OLS (xem Phụ lục 7A, Phần 7A.7 để biết kết quả in ra từ máy tính):

$$\begin{aligned}
\ln Y_i &= -3.3348 \quad + \quad 1.4988 \ln X_{2i} \quad + \quad 0.4899 \ln X_{3i} \\
&\quad\ (2.4495) \qquad\quad (0.5398) \qquad\qquad (0.1020) \\
t &= (-1.3629) \qquad (2.7765) \qquad\qquad (4.8005) \\
&\qquad\qquad R^2 = 0.8890 \qquad\qquad df = 12 \\
&\qquad\qquad R^2 = 0.8705 \qquad\qquad\qquad\qquad\qquad (7.10.4)
\end{aligned}$$

Từ phương trình (7.10.4) chúng ta thấy rằng trong khu vực nông nghiệp của Đài Loan trong giai đoạn 1958-1972 độ co dãn sản lượng của lao động và vốn là bằng 1.4988 và 0.4899. Nói một cách khác, trong giai đoạn được xem xét này, giữ nhập lượng vốn không đổi, gia tăng 1 phần trăm trong lao động dẫn đến trung bình vào khoảng 1.5 phần trăm gia tăng trong sản lượng. Tương tự như vậy, giữ nhập lượng lao động không đổi, gia tăng 1 phần trăm trong vốn dẫn đến trung bình vào khoảng 0.5 phần trăm gia tăng trong sản lượng. Cộng hai độ co dãn sản lượng lại, chúng ta được 1.9887, đây là giá trị của thông số sinh lợi theo quy mô. Rõ ràng là, trong giai đoạn này, khu vực nông nghiệp của Đài Loan có đặc điểm là sinh lợi theo quy mô tăng dần.[22]

[21] Lưu ý rằng trong hàm sản xuất Cobb-Douglas (7.10.1) chúng ta đã giới thiệu số hạng sai số ngẫu nhiên một cách đặc biệt là biến đổi nó thành dạng logarít rồi đưa về dạng tuyến tính thường dùng. Về vấn đề này, xem phần 6.8.

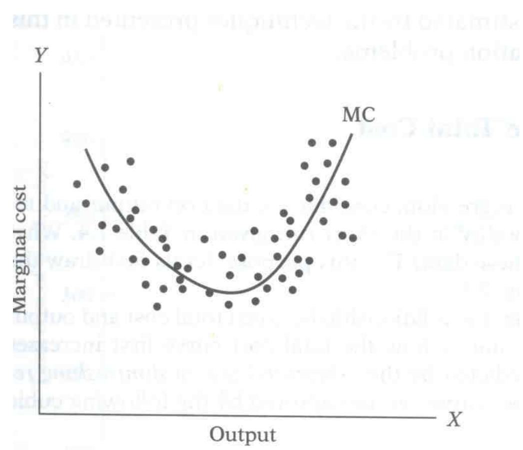
[22] Chúng ta bỏ qua câu hỏi về sự thích hợp của mô hình theo quan điểm lý thuyết cũng như câu hỏi về việc liệu người ta có thể đo lường được sinh lợi theo quy mô từ chuỗi dữ liệu theo thời gian hay không.

Từ quan điểm thống kê thuần túy, đường hồi qui ước lượng thích hợp với các dữ liệu rất tốt. Giá trị R^2 0.8890 có nghĩa là vào khoảng 89% sự biến thiên trong (logarit của) sản lượng là do (logarit của) lao động và vốn. Trong Chương 8, chúng ta sẽ thấy bằng cách nào những sai số chuẩn ước lượng này có thể được sử dụng để kiểm định giả thiết về các giá trị "đúng" của các thông số của hàm sản xuất Cobb-Douglas cho nền kinh tế Đài Loan (xem bài tập 8.15).

7.11 CÁC MÔ HÌNH HỒI QUI ĐA THỨC

Chúng ta kết thúc chương này bằng cách xem xét một nhóm các mô hình hồi qui bội, những **mô hình hồi qui đa thức**, được sử dụng rộng rãi trong các cuộc nghiên cứu kinh tế lượng có liên quan đến hàm sản xuất và chi phí. Khi giới thiệu những mô hình này, chúng tôi sẽ mở rộng thêm phạm vi của các mô hình này để có thể dễ dàng áp dụng những mô hình hồi qui tuyến tính cổ điển vào chúng.

Để xác định các ý tưởng, xem hình 7.4, hình 7.4 này cho thấy quan hệ giữa chi phí sản xuất biên tế ngắn hạn (MC) (gọi là biến Y) của một loại hàng hóa và mức sản lượng của nó (gọi là biến X). Đường cong MC vẽ cho thấy trong hình, tức đường cong hình chữ U theo sách giáo khoa, cho thấy quan hệ giữa MC và sản lượng là không tuyến tính. Nếu chúng ta phải định lượng mối quan hệ này từ các điểm rời rạc, làm cách nào chúng ta có thể thực hiện được? Nói một cách khác, loại mô hình kinh tế lượng nào thể hiện được bản chất giảm dần lúc ban đầu và sau đó tăng dần của chi phí biên tế?



HÌNH 7.4
Đường cong chi phí biên tế có dạng hình chữ U

Về mặt hình học, đường MC được mô tả trong hình 7.4 là một *parabol*. Về mặt toán học, một parabol được biểu thị bởi phương trình sau:

$$Y = \beta_0 + \beta_1 X + \beta_3 X^2 \qquad (7.11.1.)$$

được gọi là một hàm *bậc hai*, hay một cách tổng quát hơn, một *đa thức bậc hai* theo biến X- số mũ cao nhất của X biểu thị cho bậc đa thức (nếu cộng thêm X^3 vào trong hàm ở trên, nó sẽ là một đa thức bậc ba, v.v.)

Dạng ngẫu nhiên của (7.11.1) có thể được viết như sau

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \tag{7.11.2}$$

được gọi là một hồi qui *đa thức bậc hai*.

Hồi qui đa thức bậc k tổng quát có thể được viết thành

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ldots\ldots \beta_k X_i^k + u_i \tag{7.11.3}$$

Lưu ý rằng trong những loại hồi qui đa thức này chỉ có một biến giải thích ở bên vế phải nhưng nó xuất hiện với những lũy thừa khác nhau, như vậy khiến cho chúng trở thành những mô hình hồi qui bội. Nhân đây, lưu ý rằng nếu X_i được giả thiết là cố định hoặc không ngẫu nhiên, các số hạng lũy thừa của X_i cũng trở thành cố định hoặc không ngẫu nhiên.

Những mô hình này có gây ra vấn đề khó khăn đặc biệt nào về ước lượng không? Bởi vì đa thức bậc hai (7.11.2) hay đa thức bậc k (7.11.3) là tuyến tính theo các thông số β, chúng có thể được ước lượng bằng các OLS thông thường hay phương pháp ML. Nhưng còn về vấn đề cộng tuyến thì sao? Chẳng lẽ các X khác nhau này không có tương quan cao sao, bởi vì tất cả chúng đều là lũy thừa của X? Có, nhưng nhớ rằng các số hạng như X^2, X^3, X^4, v.v. đều là hàm không tuyến tính của X và vì vậy, nói một cách chặt chẽ, chúng không vi phạm giả định về phi đa cộng tuyến.[23] Tóm lại, các mô hình hồi qui đa thức có thể được ước lượng bằng các kỹ thuật đã được trình bày trong chương này và không gây ra một vấn đề mới nào về ước lượng.

Ví dụ 7.4: Ước lượng Hàm Tổng Chi phí

Để minh họa một ví dụ về hồi qui đa thức, hãy xem xét các dữ liệu trong Bảng 7.4 về sản lượng và tổng chi phí sản xuất ngắn hạn của một loại sản phẩm. Loại mô hình hồi qui nào sẽ thích hợp với các dữ liệu? Để thực hiện mục đích này, trước hết chúng ta hãy vẽ một đồ thị phân tán, như trong hình 7.5.

Bảng 7.4
Tổng chi phí (Y) và sản lượng (X)

Sản lượng	Tổng chi phí
1	193
2	226
3	240
4	244

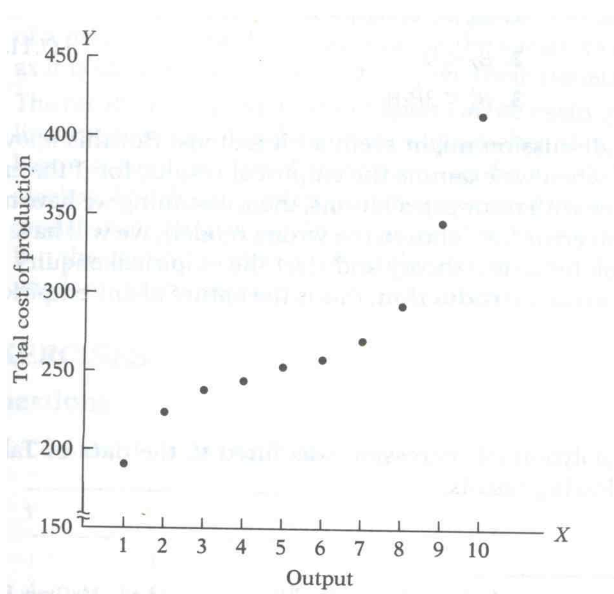
[23] Chúng ta sẽ xem vấn đề này một lần nữa ở Chương 10. Ở Chương 10 chúng ta sẽ bàn thảo một cách kỹ lưỡng tất cả các vấn đề về đa cộng tuyến.

5	257
6	260
7	274
8	297
9	350
10	420

Từ hình này rõ ràng là quan hệ giữa tổng chi phí và sản lượng tương tự như một đường cong hình chữ S kéo dài; lưu ý đường tổng chi phí lúc đầu gia tăng từ từ và sau đó tăng nhanh, như quy luật nổi tiếng về *sinh lợi giảm dần* đã dự báo. Dạng hình chữ S này của đường tổng chi phí có thể được thể hiện bởi *đa thức bậc ba* sau:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \qquad (7.11.4)$$

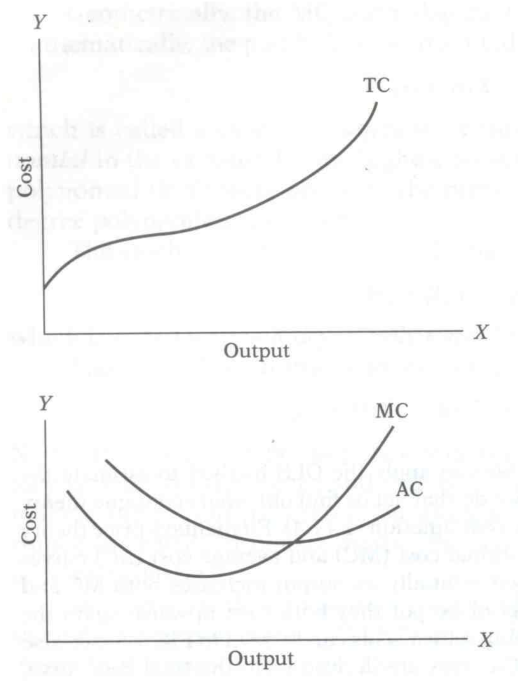
trong đó Y = tổng chi phí và X = sản lượng.



HÌNH 7.5
Đường tổng chi phí

Với các dữ liệu đã cho trong Bảng 7.4, chúng ta có thể áp dụng phương pháp OLS để ước lượng các thông số của (7.11.4). Nhưng trước khi tiến hành, chúng ta hãy tìm xem lý thuyết kinh tế nói gì về hàm chi phí bậc ba ngắn hạn (7.11.4) này. Lý thuyết về giá cơ sở cho thấy rằng các đường chi phí sản xuất biên tế ngắn hạn (MC) và chi phí sản xuất trung bình (AC) có dạng điển hình hình chữ U—ban đầu, khi sản lượng gia tăng cả MC và AC giảm, nhưng ở một mức sản lượng nhất định chúng sẽ quay lên trên, một lần nữa đây là kết quả của quy luật sinh lợi giảm dần. Có thể thấy điều này trong hình 7.6 (đồng thời xem hình 7.4). Và bởi vì các đường MC và AC là được suy ra từ đường tổng chi phí, bản chất hình chữ U của những đường này đặt một số hạn chế lên các thông số của đường tổng chi phí (7.11.4). Sự thật là có thể cho thấy rằng các các thông số của (7.11.4) phải thỏa mãn những hạn chế sau nếu chúng ta muốn quan sát các đường chi phí trung bình và chi phí biên tế ngắn hạn có dạng điển hình hình chữ U;[24]

1. β_0, β_1 và $\beta_3 > 0$
2. $\beta_2 < 0$
3. $\beta_2^2 < 3\beta_1\beta_3$ (7.11.5)



Hình 7.6
Các hàm chi phí ngắn hạn

Tất cả những bàn luận lý thuyết này có vẻ như hơi tẻ nhạt. Nhưng kiến thức này là vô cùng hữu dụng khi chúng ta xem xét các kết quả thực nghiệm, bởi vì nếu các kết quả thực nghiệm không phù hợp với các kỳ vọng tiên nghiệm, khi đó giả định rằng chúng ta chưa hề phạm một sai số đặc trưng nào (ví dụ như chọn mô hình sai), chúng ta sẽ phải sửa đổi lý thuyết của chúng ta hoặc phải đi tìm một lý thuyết mới và bắt đầu lại công cuộc truy tìm thực nghiệm lại từ đầu. Nhưng những chúng ta đã lưu ý trong **Phần Giới Thiệu**, đây là bản chất của bất kỳ một cuộc điều tra thực nghiệm nào.

[24] Xem Alpha C.Chiang, *Fundamental Methods of Mathematical Economics (Các Phương pháp Căn bản của Kinh tế học Toán học)*, ấn bản lần 3, McGraw-Hill, New York, 1984, tr.250-252.

Các Kết quả Thực nghiệm

Khi hồi qui đa thức bậc ba thích hợp với các dữ liệu của Bảng 7.4, chúng ta thu được những kết quả sau:

$$\begin{array}{lllll} Y_i = 141.7667 & + \; 63.4776X_i & -12.9615X_i^2 & + \; 0.9396X_i^3 & \\ \quad (6.3753) & \quad (4.7787) & (0.9857) & \quad (0.0591) & \qquad (7.11.6) \\ & \quad R^2 = 0.9983 & & & \end{array}$$

(Lưu ý: Các con số trong ngoặc là các sai số chuẩn ước lượng.) Mặc dù chúng ta sẽ xem xét ý nghĩa thống kê của những kết quả này trong chương sau, các bạn đọc có thể kiểm chứng rằng chúng tuân theo đúng với những kỳ vọng lý thuyết đã liệt kê trong (7.11.5). Chúng tôi để vấn đề này lại cho người đọc coi như là một bài tập để các bạn giải thích hồi qui (7.11.6).